HOT ELECTRON INJECTION EFFECT AND IMPROVED LINEARITY IN TYPE-I/II DHBT FOR MILLIMETER-WAVE MIXED SIGNAL CIRCUIT APPLICATIONS

BY

KUANG-YU CHENG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Urbana, Illinois

Doctoral Committee:

Professor Milton Feng, Chair
Professor Nick Holonyak, Jr.
Adjunct Professor Kuang-Chien Hsieh
Professor Jianming Jin

ABSTRACT

The prevalence of mobile computing devices and emerging demand for high data rate communication have motivated development of low power consumption, high performance circuits composed of compound semiconductor heterojunction bipolar transistors (HBTs).

The subject of this work is the design and fabrication of HBTs based on InP and the III-V compounds compatible with epitaxial growth on this substrate. The hot electron injection effect is incorporated to improve gain and speed by using the AlInP/GaAsSb/InP material system with a Type-I/II energy band alignment. Chapter 1 of this work gives a brief overview of the motivation to conduct this research and an introduction to other relevant work. Scaling theory, structure design and previous work done at the University of Illinois at Urbana-Champaign are presented in Chapter 2. Chapter 3 presents the submicron HBT process flows and processing challenges related to yield. The large area device results incorporated with material studies and microwave measurement of submicron devices are documented in Chapter 4. In Chapter 5, the Type-I/II HBT in this work is benchmarked and compared to foundry-manufactured Type-I InGaAs HBT. The DC and RF characterization are demonstrated. Linearity measurement is carried out to show improved high frequency distortion behaviour of Type-I/II HBT. Future work is proposed in Chapter 6.

To my family

ACKNOWLEDGMENTS

I would like to thank my adviser, Professor Milton Feng, for the opportunity to become a member of his research group and for his continuous support and guidance throughout my graduate studies. Working in a research group as accomplished as the High-Speed Integrated Circuits (HSIC) group has been a great privilege.

I am grateful for the mentoring of Professor Feng's former students Dr. Will Snodgrass and Dr. Ben Chu-Kung, both of whom demonstrated exceptional dedication and attention to detail. This work would not have been possible without the material growth expertise of Professor Keh-Yung Cheng and the hard work of his former and current students, Dr. Bing-Ruey (Barry) Wu and Chi-chih (Keith) Liao. Credit also goes to former student Dr. Michael Hattendorf and Dr. Walid Hafez for laying the foundations of the HSIC group sub-μm transistor work. I am also thankful for current and former HSIC group members Dr. Mark Stuenkel, Dr. Hanwui Then, Dr. Adam James, Dr. Wayne Wu, Huiming Xu, Eric Iverson, Rohan Bambery, Fei Tan, and Mong-Kai Wu. I have relied heavily on their support and collective knowledge of semiconductor fabrication and device measurement.

Above all, I thank my parents for their encouragement throughout the years and my friends in America and Taiwan for their support and care.

TABLE OF CONTENTS

Page

1 INTRODUCTION

In spite of the fact that highly integrated logic circuits are not feasible in either GaAs or InP substrate, the III-V material based heterojunction bipolar transistor (HBT) is indispensable in the applications in optoelectronics and wireless communications. Because of the superiority of the devices for amplifying high frequency signals, compound semiconductor HBTs are suitable for analog/ mixed-signal circuits. Progress in both material system and device processing is necessary to create transistors with power gain and current gain in the sub-millimeter wave bandwidth.

The goal of this work is the design and processing of heterojunction bipolar transistors based on InP and the III-V compounds compatible with epitaxial growth on this substrate. Based on an existing submicron HBT fabrication and layout design, the factors affecting the yield and performance uniformity are studied. In order to explore the limit of vertical scaling, variation of the material layer composition and thickness are also examined.

1.1 Development of the Heterojunction Bipolar Transistor

On December 16, 1947, J. Bardeen and W.H. Brattain demonstrated the point-contact bipolar transistor [1] while researching semiconductor materials in the Solid State Physics Group at Bell Labs. This was followed by the demonstration of the junction transistor in 1951 by Shockley, who developed the theory of minority carrier injection [2]. Commercial development of the bipolar junction transistor (BJT) soon followed.

Kroemer published the theory of using a material having a larger bandgap for the emitter of a BJT in 1957 [3], thereby forming a heterojunction at the emitter-base interface and creating the HBT. This idea was previously proposed by Shockley in a 1948 patent application as a method to increase minority carrier injection efficiency, though Kroemer developed the theory of HBT

operation independently at a time when fabricating such a device was becoming a reality. To further improve the emitter injection efficiency, gain-bandwidth and linearity (energy efficiency) of transistor operation, a single heterojunction bipolar transistor (SHBT) with a wider emitter bandgap was introduced with an AlGaAs/GaAs or InGaP/GaAs emitter, and has been used in production since 2004 as high efficiency power amplifiers in commercial handset devices, PDA, LAN, aerospace and defense products. InP based HBTs are promising to advance next generation communication system, since InP HBTs possess higher electron saturation velocity, higher breakdown voltage, lower emitter-base junction turn-on voltage (better power added efficiency), and higher thermal conductance. Since 2003, the InP SHBT has been used in 40 Gb receiver circuits and has demonstrated an increased transistor speed with current gain cutoff frequency (f_T) exceeding 850 GHz in 2006 via the use of a graded bandgap of InGaAs base and collector [4]. For mixed signal integrated circuit applications, high speed transistors are required to achieve higher breakdown voltage and linearity to improve dynamic range. The InP double heterojunction bipolar transistor (DHBT) is attractive because of the incorporation of a wider-gap InP collector to improve the breakdown voltage. Three types of InP DHBTs are reported, namely, Type-I InP DHBT (HRL, Northrop Grumman Corp., Teledyne, and Global Communication Semiconductor), Type-II DHBT (Agilent [5], UIUC [6], and ETH [7]) and Type-I/II DHBT (UIUC [8]). A Type-I DHBT has a conduction band discontinuity ($\Delta E_{C,\,BC}$ ~ 250 meV) between InGaAs base and InP collector. This base-collector interface energy barrier leads to current-blocking and gain compression effects in the collector I-V characteristics, which was also observed previously in Type-I GaAs based DHBTs with InGaP collector [9]. Hence, the Type-I DHBT requires the use of transition region (setback and superlattice layer) in the base-collector heterojunction to minimize the electron current blocking that occurs in the

collector layer. Type-II DHBT, on the other hand, has a favorable base-collector junction band alignment to curtail the blocking effect, but has a lower current gain. Reliable Type-II DHBTs with f_T and f_{MAX} = 200 GHz were developed by Agilent [10] for mixed signal circuits used in high data-rate instruments.

1.2 High Speed HBT Figures of Merit

The current-gain-cutoff frequency f_T and the power-gain-cutoff frequency f_{MAX} are two major figures of merit to characterize the high speed performance of HBTs. f_T is the frequency at which the small-signal current gain equals unity. In other words, in a common-emitter configuration f_T is the frequency where the intensity of the small signal current delivered to a short circuit load is the magnitude of the small signal current at the base input. It can be shown that unilateral current gain f_T can be decomposed to several delay terms of HBT in equations (1.1) and (1.2).

$$\frac{1}{2\pi f_T} = \tau_B + \tau_C + \tau_e + \tau_{CC} \tag{1.1}$$

$$\frac{1}{2\pi f_T} = \tau_B + \tau_C + \frac{\eta kT}{qI_C} C_{je} + (R_C + R_E + \frac{\eta kT}{qI_C}) C_{BC} \tag{1.2}$$

The base and collector transit delay τ_B and τ_C shown in both equations correspond to the transport through the epitaxial layers of the device by diffusion, drift or ballistic transport. The emitter-base and collector-base junction RC charging time constants are represented by τ_e and τ_{CC}, respectively, in equation (1.1). In equation (1.2) the two charging time terms are expressed in terms of several device parameters: η is the ideality factor used to describe the junction characteristics commonly shown in a Gummel-Poon plot. The value can vary between 1 and 2 depending on layer structure design and process conditions. $\eta kT / qIc$ is the dynamic emitter junction resistance (also represented by r_e). C_{je} is the emitter junction capacitances. R_C is the collector resistance and R_E is the emitter resistance. C_{BC} is the base collector depletion

capacitance. It is obvious that C_{BC} charges through both dynamic, extrinsic emitter resistances and collector resistances. From these equations, basically HBTs can demonstrate faster speed under higher current density (the $1/I_C$ dependence).

Equation (1.3) is the expression of f_{MAX}, the unilateral power gain, also known as the maximum frequency of oscillation. f_{MAX} is not only related to f_T but also dependent on the base resistance R_B and base collector capacitance, as shown in (1.3).

$$f_{\mathrm{MAX}} = \sqrt{\frac{f_t}{8\pi\ R_B C_{BC}}} \tag{1.3}$$

The most effective way to increase the cutoff frequency of the HBT is through vertical scaling of the base and collector layer to reduce the intrinsic forward delay of the device. It has been verified from earlier generations of HBTs processed at UIUC that greater than 75% of the total delay was associated with the transport delay. To reduce total charging time, on the other hand, layer doping level and bandgap engineering are involved. The C_{je} can be lower with lowly doped emitter layer. Emitter resistance R_E and collector resistance R_C can be reduced by using a highly doped and narrow bandgap cap layer for low contact resistance. Although vertical scaling can improve frequency performance directly and effectively, there are many drawbacks at the same time. As base and collector layers are thinned, parasitic resistances and capacitances also increase. The breakdown voltage is also lower, which is not favorable in future circuit applications. As can be seen in equation (1.3), the increase in R_B and C_{BC} will always lead to a reduction of f_{MAX}, yielding a critical f_T / f_{MAX} tradeoff design parameter.

1.3 Velocity Overshoot and Ballistic Transport

In 1972 Ruch [11] first noticed velocity overshoot in Monte Carlo simulations of electron transport. Velocity overshoot is a term applied to situations where, in the presence of an electric

field, the average electron velocity is greater than what would be obtained in steady state with a spatially constant electric field in a bulk semiconductor. Velocity overshoot can happen when an electric field changes rapidly in time or space or both, and the field is large compared to the field where electron velocity saturates. This phenomenon is associated with the finite time required for an ensemble of electrons in an electric field to redistribute into nonequilibrium, steady-state distribution. The electron velocity versus electric field strength curve for Si increases with field strength before saturating at a velocity slightly less than 10^7 cm/s. The direct-gap compound semiconductor materials, however, exhibit a negative differential mobility behavior. In the compound semiconductor materials, the electron velocity peaks at $2\text{-}3\times10^7$ cm/s at a low field intensity before decreasing to a saturation velocity at $1\text{-}2\times10^7$ cm/s at higher field intensities.

In a typical direct bandgap compound semiconductor material's *E-k* diagram, the conduction band minimum is the Γ-valley, which exhibits a low effective mass. The next-highest energy valley is the *L*-valley, having a larger effective mass. Upon a spatially or temporally sudden application of a large electric field in these materials, electrons will accelerate (subject to intervalley scattering) in the low effective mass, high velocity Γ-valley until they are scattered into the larger mass, low velocity *L*-valley. After a long time, most of the electrons will be in the L-valley because of its larger density of states. However, for some short interval of time, a certain number electrons can achieve large velocities in the Γ-valley before discrete scattering events redistribute to the L-valley, and a significant velocity overshoot can be expected. The velocity overshoot phenomenon can be utilized to improve the carrier transport behavior in HBTs and enhance the operation frequency. When electrons transport through base and collector, if they acquire high enough energy to keep high velocity but not so much energy that they scatter to the L-valley and slow down, average electron velocities exceeding saturation

velocity can be achieved. As a result, the collector delay time can be reduced and we can get higher f_T and f_{MAX}.

Velocity overshoot is related to, and often confused with, ballistic transport. The term *ballistic transport* implies electron drift or acceleration in the presence of an electric filed, such that the motion is unperturbed by scattering events. At a Type-I alignment emitter-base heterojunction where the conduction band of the wide-gap emitter is positioned above the conduction band of the base, the injected electrons will have an initial kinetic energy equal to the value of the discontinuity. This is known as *hot electron injection*. The ballistic transport only occurs over very short distances until the electron is thermalized due to scattering [12], [13], [14]; however, the residual energy enhances the electron migration.

1.4 Materials Selection in InP-Based HBT

Compound semiconductor materials with lattice constants near that of InP are candidates for use in InP based HBTs. The difference in energy gap and the alignment of two materials' energy bands dictates their usefulness in designing the HBT. The two most commonly used band alignment systems are InP/InGaAs, which is a Type-I junction, and InP/GaAsSb, which is Type-II. Both of these junctions present a significant barrier to reverse injection of holes from the narrow gap material to the InP and are thus good for forming the emitter-base junction in an n-p-n HBT. The Type-I alignment of the InP/InGaAs causes electrons injected from an InP emitter to an InGaAs base layer to possess an initial kinetic energy that can lower the base transit delay in an HBT. The Type-II alignment of InP/GaAsSb results in purely thermal injection of electrons from an InP emitter to a GaAsSb base. In an SHBT, the narrow-gap base material is also used for the collector region. It order to enhance the breakdown voltage and improve velocity overshoot in the collector depletion layer, DHBTs with a wide gap material used for the

collector are preferred. However, the base-collector junction is critical and needs extra care to avoid introducing a carrier transport barrier. The Type-II GaAsSb/InP junction is useful for DHBT design because it favors electron collection when used as a base-collector material.

2 SUBMICRON HBT SCALING

2.1 Scaling Theory and Layer Design

An HBT has vertical structure in terms of epitaxial growth. The fabrication basically follows a top-down mesa etch process followed by metal contact deposition to define the device dimensions and make electrical contact to the emitter, base, and collector layers. The nonplanar architecture leads to a lot of tradeoffs in the fabrication at the device level that are just as important to the final performance as the original design of the material. In general, while transit delays are reduced through vertical scaling, the RC charging times must be reduced through lateral scaling— both reducing junction areas and bringing metal contacts closer to the intrinsic region of the device.

The emitter width determines the active device area. Ignoring current spreading that occurs in base and collector layers, the regions to both sides of the area underneath the emitter contact are external. The parasitic resistance and capacitance due to external area of the device are thus defined by the emitter width. The current from these contacts must flow laterally through the base and subcollector regions, respectively, to reach the active region because the base and collector contacts are placed in the external region. The laterally feeding resistance changes the base-emitter potential in the extrinsic base region (from base contact to emitter mesa). The correlated effect is known as emitter current crowding, which happens around the edge of the emitter due to the potential difference.

The base sheet resistance is usually high in thin base HBTs such as those presented in this work. The high current density operation also makes the current crowding significant. The current crowding effects together with high base resistance seriously degrade the performance of wide emitter geometries, as the parasitics associated with the large device are present. The

emitter contact resistance, which is approximated by the emitter specific contact resistance divided by the contact area, sets the limit to emitter scaling. For current structures, the emitter resistance significantly degrades performance for emitter widths less than 0.2 μm.

In order to reduce the base resistance, the base is heavily doped to lower the resistivity. Also, the separation between the base contact and the emitter (S_{BE}) is minimized by using self-aligned base contact. During the emitter etch, S_{BE} is then controlled by the wet chemical etch undercut of the emitter. The base contact width is the width of the deposited base contact minus any undercutting of the metal during the base-collector mesa etch. Undercutting the base contact will reduce the base-collector area and thus C_{BC}. However, the base contact resistance R_{Bc} and total base resistance will then increase, especially as the contact width falls below the ohmic contact transfer length of the base layer. The control of the base-collector mesa etch undercut can determine the f_T/f_{MAX} tradeoff parameter because f_{MAX} is dependent on both R_B and C_{BC}, while f_T depends only on C_{BC}.

The collector resistance components are similar to those of the base, with a component due to the contact, a component due to lateral conduction from the contact to the device edge (S_{BC}), and a spreading term under the device. The collector resistance is dominated by the contact resistance and the total collector resistance is small compared to the emitter and dynamic emitter-base junction resistances that also contribute to the collector junction charging delay.

In addition to the layout scaling theory addressed above, layer structure design is also critical to high speed HBT design. The fundamental working theory of the HBT is to assemble a wide gap emitter with a narrow gap base to prevent backwards minority carrier injection from the base to the emitter. In the InP-based material system we are interested in, the emitter material is chosen to be InP/InAlP, which has a large valence band discontinuity in regard to the InGaAs or

GaAsSb base layer. The advantage of HBT over BJT lies in the fact that HBT can have high injection efficiency by manipulating the band offset while BJT requires control over the ratio of the emitter to base layer doping concentration. Low doping throughout the full emitter layer would lead to substantial ohmic resistance, while higher doping causes larger depletion capacitance at the emitter-base junction. The former increases both R_E and r_e and the latter will increase the parasitic capacitance C_{je}, both leading to reduced high frequency performance. To overcome this problem, a highly doped n-type cap layer is grown on top of the lowly doped region near the emitter-base junction. Narrow bandgap materials such as InAs or InGaAs are used for the cap to build a low resistance ohmic contact. The dynamic resistance of the forward-biased junction will be affected by the conduction band offset of the emitter-base heterojunction.

As for the base layer design, a narrow gap base region is chosen with band alignment to the emitter material that prevents majority carrier flow out of the base. To be lattice matched to InP, InGaAs is the most common base material in InP-based HBT. GaAsSb is, however, favored in this work because its band alignment forms Type-II alignment with InP, enabling the use of InP as the collector material. To get high power gain, the p-type base is highly doped to lower the lateral and ohmic contact resistance. In modern HBT design the very thin base suffers degraded f_{MAX} due to high base resistance, although reduced base doping can enhance base transport and hence current gain by reducing recombination and scattering event. A thicker base layer will result in lower base resistance, while a thinner base layer reduces the base transit delay—creating a pivot point in designing the f_T / f_{MAX} balance of a transistor. Aggressive parasitic reduction is necessary to maintain f_{MAX}, as transistor structures are vertically scaled to improve f_T. We can solve this tradeoff by implementing a compositional grading in the base layer to vary the energy gap and create a built-in electric field to promote forward transport of minority carriers.

Compared to a uniform base where carriers travel by diffusion, the graded base (either by compositional or doping grading) with same thickness can significantly reduce the transit time due to the built-in field. Because the doping has remained constant, there is no base resistance penalty as there is with vertical scaling of the base, and thus both f_T and f_{MAX} are improved.

To accelerate carrier transport with a favorable band alignment, the collector material should have a conduction energy band edge lower than that of the base material and a high mobility to increase the velocity in the depletion region.

A wide gap material will have higher breakdown field strength, giving the transistor a higher breakdown voltage. An InP collector paired to a GaAsSb base meets all of these criteria. The thickness of the collector depletion region directly determines the base-collector capacitance C_{BC} and the collector signal delay τ_C. Vertical scaling to reduce τ_C will improve f_T as long as signal delay is dominant. Vertical scaling fails once the increase in C_{BC} from vertical scaling makes charging delays comparable to transit delay.

At high current densities, the mobile charge concentration in the collector can rise to levels comparable to that of the collector doping. At this point, the electron charge begins to compensate for the positive charge of the uncovered donors and induces excessive holes in order to maintain quasi-neutrality. The electric field at the base-collector junction will start to decrease and the high-filed region, originally located at the physical base-colector junctions, is relocated. If the base-collector is a homojunction, or if the valence band discontinuity is small, the region of the collector with excess holes becomes an extension of the p-type base, effectively extending the base region. This "base push-out," also known as the Kirk effect [15], will increase the base transit time, increase C_{BC}, and lower current gain. If the base-collector junction is a heterojunction with a significant barrier for holes, the built-in barrier will prevent base push-out,

but electrical field relocation can still occur. The large mobile electron charge will both reduce the field in the collector and induce hole buildup in the base near the collector, thereby raising the conduction band edge in both the base and collector and increasing electron transit times. These effects can be observed in the DC common-emitter family of curves as current gain compression at high current densities. In addition, the device will typically require a larger reverse bias at the base-collector junction to reach peak current gain and this will manifest as a progression of the knee voltage to larger values with increasing current density that is more significant than the typical dependence due to emitter and collector resistances. These effects lead to premature decline of the cutoff frequency of the device versus current density—limiting the high frequency performance of the device.

2.2 Vertical Scaling of SHBTs

Vertical scaling has previously been applied to InP/InGaAs SHBTs by Hafez et al. at the University of Illinois [16]. Vertical scaling and compositional grading reduced the traditionally dominant electron transport delay, resulting in record f_T as high as 765 GHz [17]. The vertical scaling reaches a limit, however, as parasitic charging delay rises when collector thickness is $\leq$ 100 nm. The impact of rising capacitance with vertical scaling on f_T is partially offset by increasing the operating current density to reduce emitter charging delay and by reducing parasitic extrinsic device area, both of which have helped keep total charging delays approximately constant with vertical scaling.

2.3 Type-II InP/InGaAsSb-GaAsSb DHBTs

Bolognesi et al. of Simon Frasier University began significant work on GaAsSb base devices in 2001 [18], verifying the Type-II alignment of GaAsSb to InP via photoluminescence and eventually demonstrating excellent RF performance with f_T and f_{MAX} of 300 GHz [19] with f_T

increasing to 384 GHz after vertical scaling of the base and collector [20]. Even after vertical scaling, the cutoff frequencies of these devices trailed those of SHBTs and Type-I DHBTs because of low electron diffusivity in the GaAsSb base, which has lower mobility than InGaAs, and no ballistic injection from the emitter to base because of the Type-II alignment of GaAsSb to InP. Snodgrass et al. [21] at UIUC have demonstrated for the first time the incorporation of a compositionally graded base layer (InGaAsSb-GaAsSb) into Type-II DHBTs to take advantage of the aforementioned inherent benefits of the GaAsSb/InP material system using the same scaling principles previously applied to InP/InGaAs SHBTs at Illinois [16]. Vertical scaling of the base and collector layers is also employed, with parasitic resistances and capacitances controlled via lateral scaling. Figure 2.1 presents the first reported Type-II DHBT to use a graded base, 200 Å base and 650 Å collector to achieve record high f_T exceeding 600 GHz.

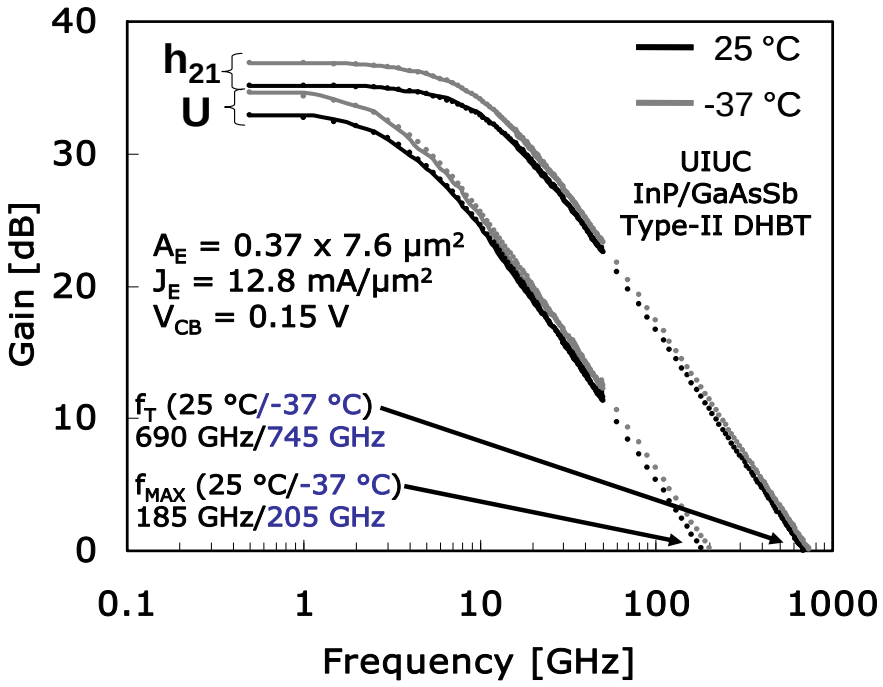

Figure 2.1. RF measurement of UIUC submicron Type-II DHBTs with 200 Å base-650 Å collector.

2.4 Design of Type-I/II DHBTs

To boost f_T and f_{MAX} simultaneously in order for balanced analog and mixed signal design, further improvement of the base grading scheme is evaluated as well. High base doping is favorable to lower base sheet resistance, but the increase of scattering also accelerates carrier recombination and lowers the current gain. To counteract this effect, we take advantage of hot electron injection of an abrupt emitter-base heterojunction where the conduction band of the wide-gap emitter is positioned above the conduction band of the base. It has been demonstrated that the "emitter launcher" can improve dc current gain. This way we can advance the performance of Type-II InP/GaAsSb DHBTs to exploit the scaling potential. Other researchers have recently used similar approaches to achieve higher current gain and higher f_T in Type-II DHBTs [22]-[25]. The InGaP emitter could be used to achieve a near zero offset junction with ~15% Ga composition and ~1.0 % lattice mismatch. A higher Ga concentration could be used to achieve a Type-I offset, but the strain would be significant and the structure difficult to grow without defects. InAlP as the emitter is attractive because only a small Al concentration is required to raise the emitter enough to achieve the type-I alignment so that the strain is minimal. In addition, the base can also be compositionally graded using a small Al concentration gradient to achieve a significant built-in field. Though InAlAs has a type-I alignment to GaAsSb at their lattice-matched compositions, the high 48% Al concentration of the InAlAs creates some processing challenges, as the wet etch for the material is not completely selective to the base layer. The InAlP emitter combined with an GaAsSb graded base is the most likely to succeed, and the growth of this interface has already been demonstrated by both MOCVD and MBE ([25], [26]). Figure 2.2 presents an energy band diagram of the first structure studied as part of this work. The structure, which was the first reported Type-I/II DHBT to use a graded base, has a

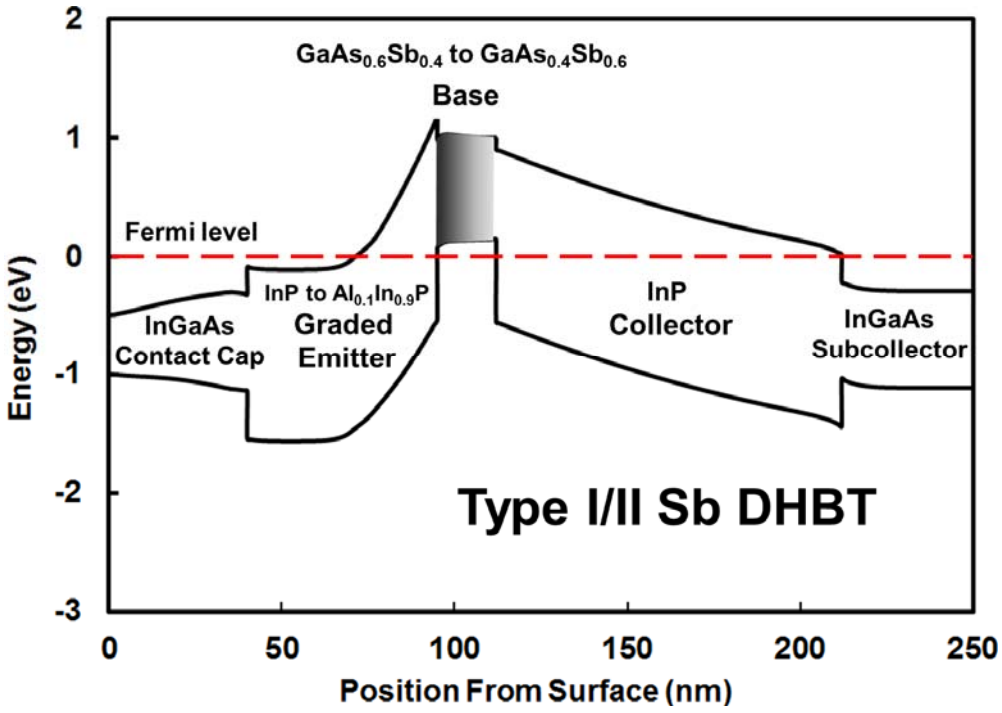

Figure 2.2. Equilibrium energy band diagram of a UIUC Type-I/ Type-II graded-base DHBT with 170 Å base and 1000 Å collector.

170 Å $GaAs_xSb_{1-x}$ graded base layer and 1000 Å InP collector. The emitter-base junction band offsets are calculated based on the model-solid theory [27] and include band edge shifts due to strain for the materials not lattice matched to InP. The epitaxial structures for these devices were grown by molecular-beam epitaxy (MBE) on <100> InP:Fe semi-insulating substrates. The structure consists of a 200 nm InGaAs subcollector doped to $n = 5 \times 10^{19}$ cm^{-3}, 100 nm InP collector lightly doped to $n = 5 \times 10^{16}$ cm^{-3} to achieve a high breakdown voltage, 17 nm carbon-doped GaAsSb base (8×10^{19} cm^{-3}, $R_{SB} = 2000$ Ω/sq), AlInP emitter graded to InP with Si doped to 3×10^{18} cm^{-3}, and a 40 nm compositionally graded emitter cap ($In_{0.53}Ga_{0.47}As$ to InAs) doped to $n = 5 \times 10^{19}$ cm^{-3}. The emitter is linearly graded over 15 nm from InP to $Al_{0.1}In_{0.9}P$ at the emitter/base interface to create an energy launcher for hot electron injection to improve current gain and speed. The abrupt emitter has a conduction band offset equal to 140 meV relative to the

base, as illustrated by the emitter-base junction of the graded structure in Figure 2.2. An aluminum concentration of 10% is chosen as the emitter ramp to minimize the strain induced from lattice mismatch and control the junction quality. To generate a built-in quasi-electric field in the base, in order to accelerate the carrier transport, base grading is implemented into the layer structure. The composition of the 17 nm graded base is $GaAs_{0.4}Sb_{0.6}$ at the base-collector junction and is graded to $GaAs_{0.6}Sb_{0.4}$ at the emitter-base junction. The energy gap variation across the compositionally graded base layer, calculated using the Model Solid Theory, is approximately 20 meV [28]. The Type-II conduction band lineup between the base and the collector not only eliminates the current blocking effect observed in Type-I InGaAs/InP DHBT, but also provides hot electron injection from base through collector. A conduction band discontinuity of approximately 50 meV remains at the base-collector $GaAs_{0.4}Sb_{0.6}$/InP interface.

3 PROCESSING TECHNOLOGY

3.1 Submicron HBT Fabrication

Figure 3.1 describes the process flow for the HBT fabrication process used in this work. Figure 3.2 presents schematic images of an HBT after key fabrication steps have been completed. The process begins by deposition of the submicron width, hexagonal emitters and square contact posts for the base and collector. These are defined by electron beam lithography (EBL). This step is shown in the top section of Figure 3.2. The emitter metal is used as the etch mask for a wet chemical emitter etch. Following the emitter etch, the base metal pattern is defined by EBL and the thin base metal is deposited. This point in the process is depicted by the middle section of Figure 3.2. Following the base metal deposition, silicon nitride is deposited on the sample via plasma-enhanced chemical vapor deposition (PECVD). The silicon nitride will protect the emitter mesa during the base-collector wet chemical etch. The formation of the sidewall spacer is done by self-aligned reactive ion etching (RIE). After the field silicon nitride is removed, the base-collector etch will be performed. The silicon nitride is then etched-back to expose the emitter metal. Collector metal is then defined by EBL and deposited. After collector metal deposition, an isolation etch mask is defined by optical lithography. This pattern protects the intrinsic device and leaves only the base contact post and base air-bridge exposed for etching. The isolation etch undercuts the base air-bridge to remove parasitics associated with the base post, and also etches down to the semi-insulating substrate around each device to isolate the transistors from one another. The isolation step is shown by Figure 3.3(a), which highlights the back-end process. Figure 3.3(a) clearly shows the material removal from underneath the base metal air-bridge. The isolated devices are then planarized in bizbenzocyclobutene (BCB), which is etched-back to expose the three contact posts as shown in Figure 3.3(b). Finally, the RF probe

pads are defined via optical lithography and the metal is deposited on the BCB, contacting the exposed base, emitter, and collector posts. This final step is shown in Figure 3.3(c), where the device is visible through the BCB, though only the contact posts are exposed. Fabricated emitter metal widths range from 0.35 μm to 0.55 μm, with the emitter material subsequently undercut approximately 0.05 μm on each side yielding fabricated emitter widths from 0.25 μm to 0.45 μm. Total self-aligned base contact deposition widths range from 0.65 μm to 1.1 μm. This yields individual base contact widths on either side of the emitter ranging from approximately 0.2 μm to 0.35 μm which are undercut during the base-collector etch to as small as 0.1 μm.

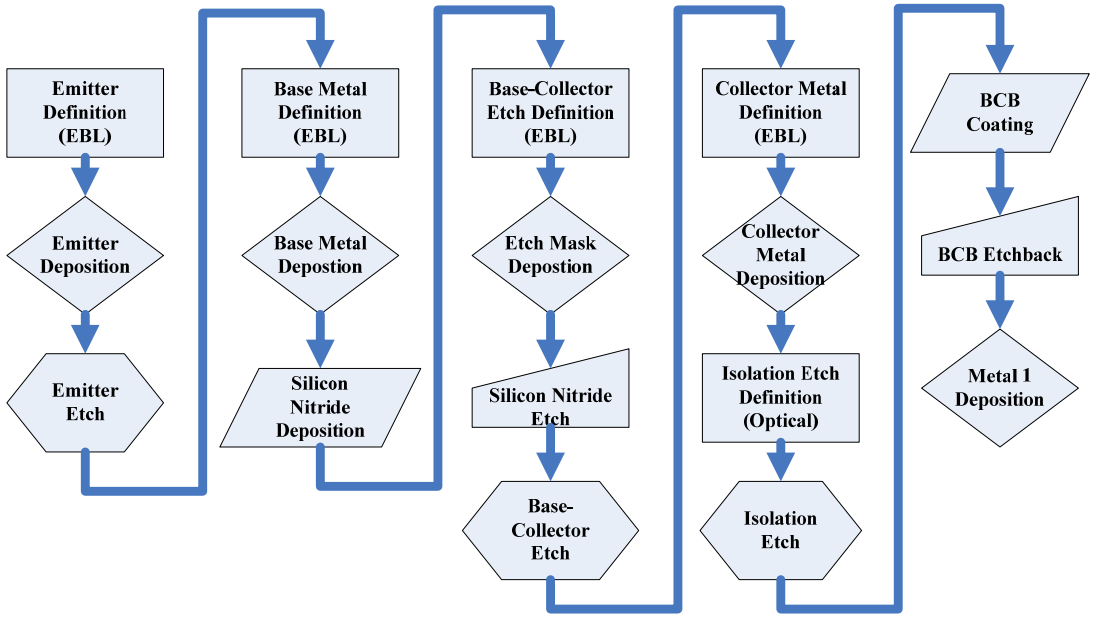

Figure 3.1. UIUC submicron HBT process flow.

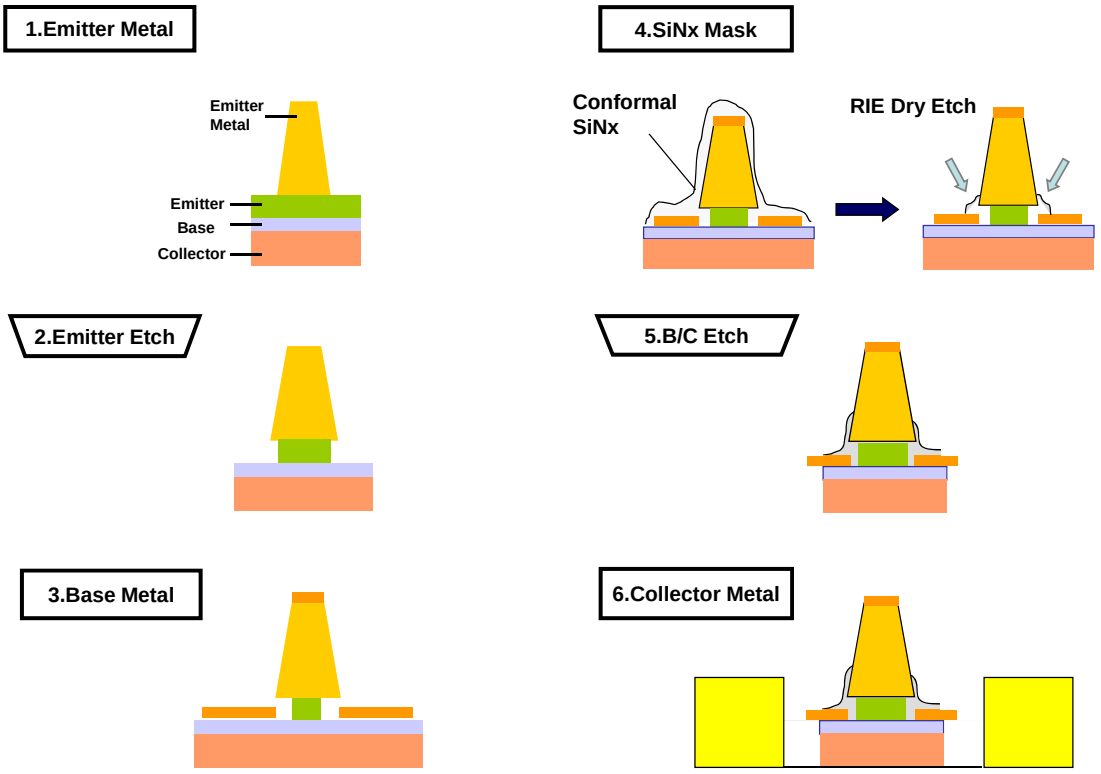

Figure 3.2. Cross section schematic views after the emitter deposition, emitter etch, base deposition, base-collector etch, and collector deposition.

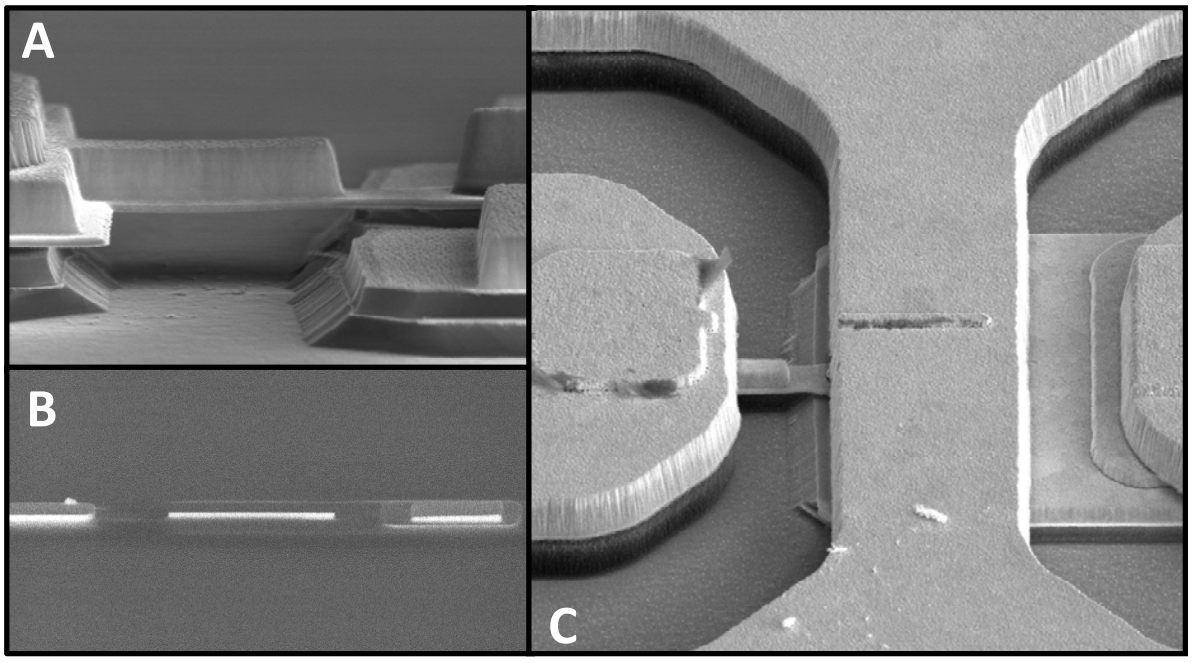

Figure 3.3. SEM images of the back-end process showing the device after (a) isolation etch, (b) BCB planarization and etchback, and (c) probe pad deposition.

3.2 Process Issues

The success of submicron HBT fabrication requires careful control of chemicals, material characterization, and equipment maintenance. It was noted that the electron beam evaporator used to deposit metal contact is especially critical to yield the desired contact profile and succeeding wet etch process. Since emitter metallization is the first step to define the active device features, extra care needs to be taken.

3.2.1 Emitter Metallization and Self-Aligned Etch

As described in section 3.1, the submicron emitter metal is defined by electron beam lithography. The photoresist is composed of a double-layer stack with an electron beam sensitive top layer to define the pattern opening size and a bottom layer to create an overhang to facilitate the liftoff process. Figure 3.4(a) shows the schematic of the bi-layer resist. During the metal evaporation, the standard emitter metal stack (15 nm of titanium, 15 nm of platinum, and 500 nm of gold) is evaporated in order. The chamber is brought to high vacuum. However, the vapor pressure of each metal is high enough that during the evaporation, metal molecules can migrate freely on the surface. As a result, the actual metal line width is wider than the top photoresist opening, creating an unwanted thin "wing" around the periphery of the emitter metal (shown in Figure 3.4(b)). The succeeding wet etch of the emitter mesa relies on the emitter metal as mask. If the emitter wing extends too far or the structure is too thick, then after the emitter etch the metal could collapse to impede further undercut etch, as Figure 3.4(c) shows.

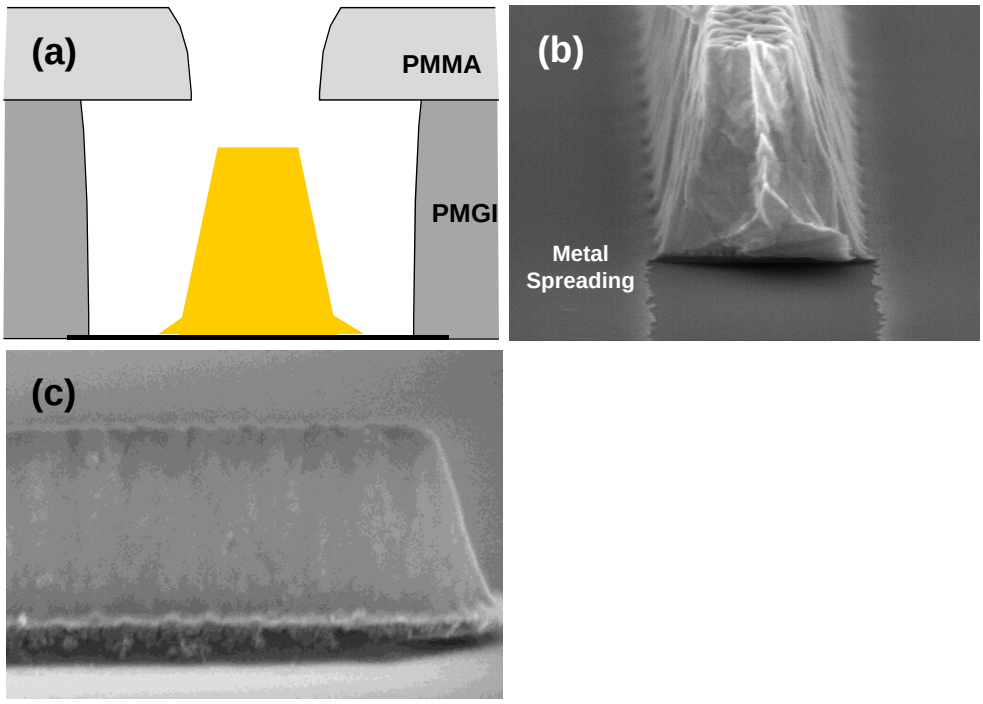

Figure 3.4. (a) Bilayer photoresist profile of emitter metal, (b) metal "wing" spreading during evaporation, and (c) metal collapse during wet etch.

The collapsed wing folds around the emitter metal, with the detrimental consequence that the base and emitter metals are shorted. This is due to the fact that base metal deposit is self-aligned and the undercut gap prevents the emitter from shorting to the base. It is apparent that the folded metal fills the gap and creates the metal connection.

To address this critical issue, we transfer the process from the original evaporator to a newly acquired one. The new evaporator was optimized in terms of heat shielding and a temperature probe was installed close to the sample holder. Figure 3.5 shows a photograph of the whole apparatus and a temperature record of emitter metal evaporation. The reduced thermal cycle compared to the old evaporator indicates the photoresist pull-back is mitigated so the possible metal spreading is minimized.

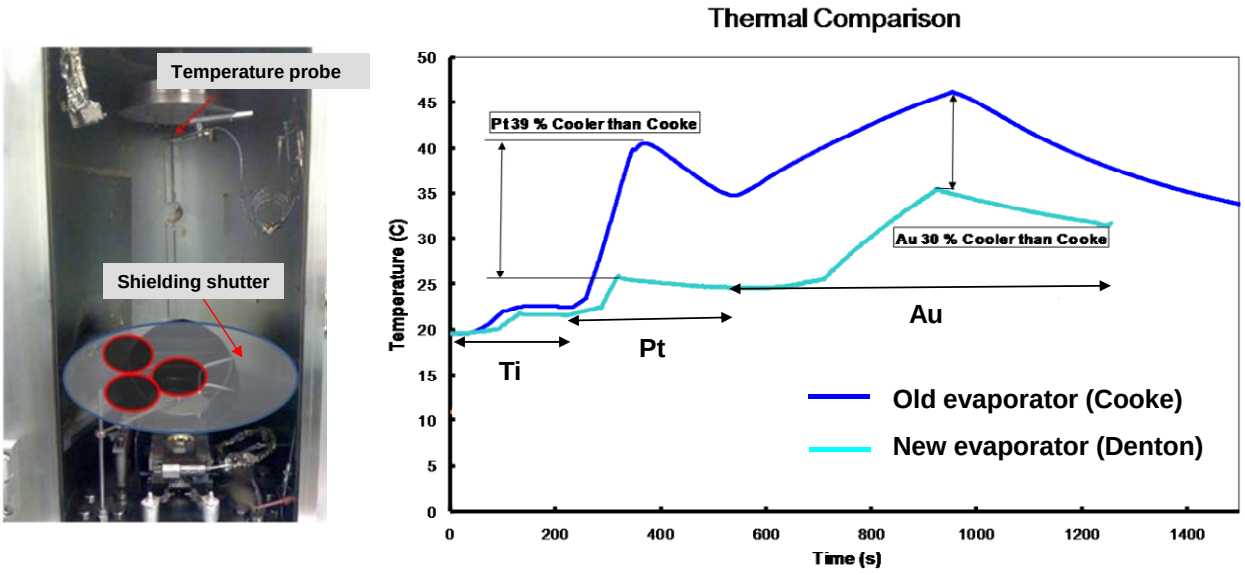

Figure 3.5. The new evaporator photograph (left) and the temperature record in two evaporators showing reduced thermal effect in the optimized evaporator.

Not only the evaporator itself but also the photoresist stack and metal crucible size play important roles in the optimization work. As Figure 3.6 (a) illustrates, the larger the source, the farther the metal spreading. Also, the photoresist thickness can be seen as another factor determining metal spreading. This solid angle theory helps to understand the root cause of the metal-wing formation.

By replacing a 1.25 cm diameter gold crucible with a 0.5 cm one, and replacing a 6000 Å thick photoresist with a 2500 Å thick one, the wing size was finally reduced from 75 nm to 25 nm at each side of the emitter (Figure 3.7). The shirked wings prevent base-to-emitter short in succeeding fabrication steps.

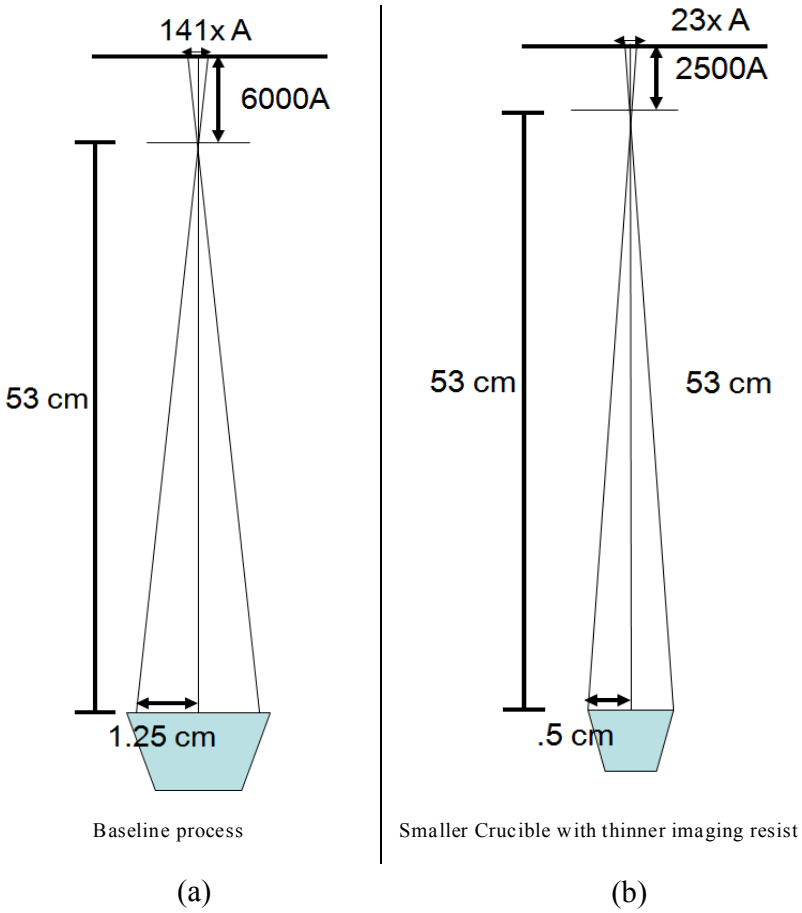

Figure 3.6. (a) Geometric factors affect evaporated metal spreading foot length. (b) Using smaller metal crucible and thinned photoresist shrinks the spreading.

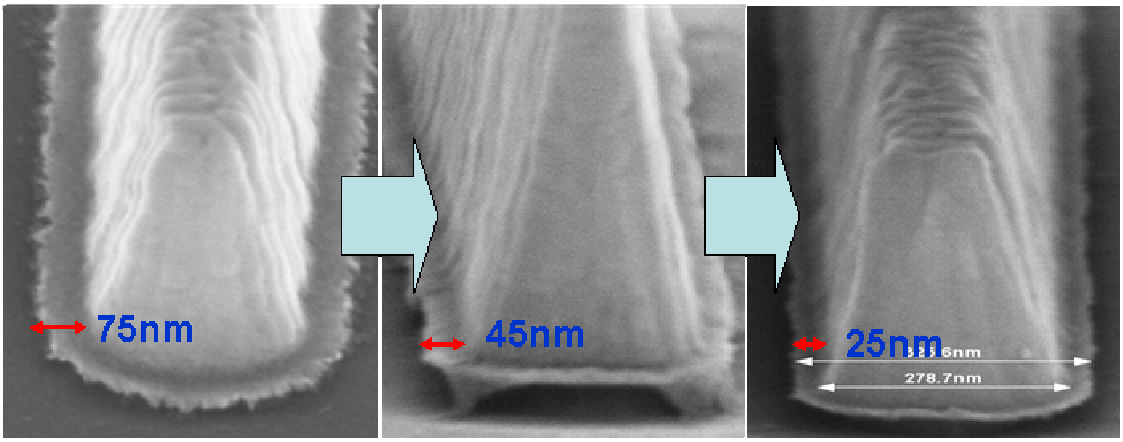

Figure 3.7. The evolution of emitter wing size reductions done by optimization of evaporator and photoresist stack.

3.2.2 Device Isolation Etch

Submicron HBTs are designed for high speed operation. This means all parasitic components need to be eliminated. The material underneath the base contact posts can be considered as parasitic base-collector capacitance and should be disconnected from the active device. The base post is connected to the device with the air-bridge, which is a rectangular metal about 1 um wide. During the isolation etch the device will be covered with photoresist and the air-bridge is exposed. Depending on the material structure, the DHBTs may have a more or less complicated etching procedure. Basically, different acids need to be alternatively applied to fully clear the material underneath the base air-bridge. The procedure begins with a citric etch to remove most of the InGaAs subcollector cap layer, followed by a quick dip in HCl to partially etch away InP (collector and subcollector layer). It is important to know that InP has a specific crystal plane etching preference and basically the etchant cannot isotropically etch InP. In the case of HBT isolation, this indicates that any layer above InP has to be completely removed so that chemicals can attack the InP from every direction. This is the reason why the first InP etch cannot be long enough to remove all the material. In between each etching, it is also recommended to use SEM to check if the air-bridge is clear. Before the epitaxial layer growth, the growers usually grow a buffer layer (same material as the substrate) on top of the substrate to begin the epitaxial growth with a smoother surface. However, the background doping concentration of the furnace actually makes the buffer layer conductive. It is necessary to completely remove the buffer layer to isolate the transistors from one another.

4 DEVICE RESULTS

4.1 Large Area Device Data

Two material structures studied in this work are listed in Table 4.1. The Type-I emitter-base heterojunction alignment utilizes strain $In_{(1-x)}Al_{(x)}P$ as emitter material with $GaAs_{(x)}Sb_{(1-x)}$ as base material. A graded emitter cap is incorporated to reduce contact resistance.

Structure GS1493 is designed to target a high f_T by using a thin base and collector for low transit delay. GS1494, on the other hand, is designed to achieve simultaneously high f_T and f_{MAX} by using thicker base layers, increased base doping concentration for lower base resistance and a thicker 1200 Å collector to lower C_{BC}.

Table 4.1. Material structures studied in this work.

GS 1493	Material	x	Thickness (Å)	Dopant	Level (/cm^3)	Type
Cap	In(x)Ga(1-x)As	0.53 to 1	350	Si	Maximum	N+
Cap	In(x)Ga(1-x)As	0.53	50	Si	Maximum	N+
Emitter	InP		300	Si	Maximum	N+
Emitter	InP		75	Si	3.0E+18	N
Emitter	In(1-x)Al(x)P	.1 to 0	175	Si	7.0E+17	N
Base	GaAs(x)Sb(1-x)	0.4 to 0.6	170	C	7.0E+19	P+
Collector	InP		650	Si	5.0E+16	N
Subcollector	In(x)Ga(1-x)As	0.53	650	Si	Maximum	N+
Subcollector	InP		3,500	Si	Maximum	N+
Etch Stop	In(x)Ga(1-x)As	0.53	100		UID	
Substrate	InP					
GS1494	**Material**	**x**	**Thickness (Å)**	**Dopant**	**Level (/cm^3)**	**Type**
Cap	In(x)Ga(1-x)As	0.53 to 1	350	Si	Maximum	N+
Cap	In(x)Ga(1-x)As	0.53	50	Si	Maximum	N+
Emitter	InP		300	Si	Maximum	N+
Emitter	InP		75	Si	3.0E+18	N
Emitter	In(1-x)Al(x)P	.1 to 0	175	Si	7.0E+17	N
Base	GaAs(x)Sb(1-x)	0.4 to 0.6	300	C	7.0E+19	P+
Collector	InP		1,200	Si	5.0E+16	N
Subcollector	In(x)Ga(1-x)As	0.53	650	Si	Maximum	N+
Subcollector	InP		3,500	Si	Maximum	N+
Etch Stop	In(x)Ga(1-x)As	0.53	100		UID	
Substrate	InP					

The structures studied in this work were first characterized by DC measurements to characterize the resistances of the emitter, base, and collector contact layers, as well as the transistor action of large area 50×50 μm^2 emitter transistors. Sheet and contact resistances were measured using the transfer-length method (TLM) [29]. The TLM pattern uses 60-μm pads

placed on isolated mesas at spacings varying from 30 μm to 5 μm. The large area transistors have emitter dimensions 50 μm × 50 μm with base contact to emitter spacing and collector contact to base-collector mesa spacings of 5 μm. DC measurements were taken using an HP4145B Semiconductor Parameter Analyzer and Cascade Microtech 12-μm DC probe tips. Two probes per pad were used in a Kelvin arrangement during TLM measurements to subtract the effects of cable and probe resistance from the measured data.

The base layer is the most challenging part of the HBT epi-growth because usually the base layer is thin (200 Å to 300 Å). And the graded base requires a precise control of different sources in a very short time. The grading profile, base doping concentration, and junction integrity all have great influence on the device performance. The base sheet resistance is a critical parameter for f_{MAX} performance. Figure 4.1 plots the measured base sheet resistance for each of the structures versus their base layer thickness. The two structures in this work, GS1493 and GS1494, are compared to previous work on Type-II alignment DHBTs material, GS620 and GS637. GS1493 and GS620 have similar base and collector layer thickness. GS637 and GS1494 also have the same base layer thickness. This makes it easier to observe the influence of inserting the emitter launcher (strained InAlP) on enhancing current gain. Figure 4.1 shows that in order to achieve higher f_{MAX}, the doping concentration is raised compared to previous materials. As a result, for 300 Å base thickness, GS1494 has base resistance 50% less than GS637. Similarly, in GS1493, which has 170 Å of base layer, the base sheet resistance is slightly lower than GS620, whose base layer is 30 Å thicker than GS1493. An expected advantage of the graded base is that thicker base layers can be implemented for smaller resistance with only a minimal transit time increase if electrons are accelerated to high velocities in the first half of the base region.

Current gain is expected to decrease with increasing base thickness as the transit time grows. It typically also decreases with decreasing sheet resistance because higher impurity concentrations lead to shorter minority carrier recombination lifetimes. The observed current gain vs. base thickness is plotted in Figure 4.2. It is observed that for 300 Å thick base layer structures, GS1494 has a lower current gain than GS637. This corresponds to the lower sheet resistance due to higher doping in GS1494.

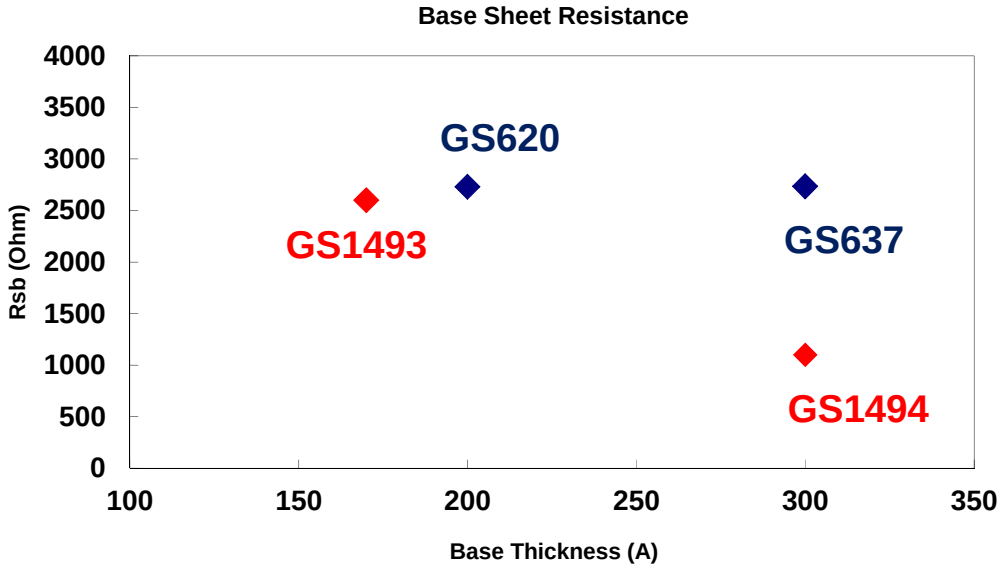

Figure 4.1. Base layer sheet resistance versus layer thickness.

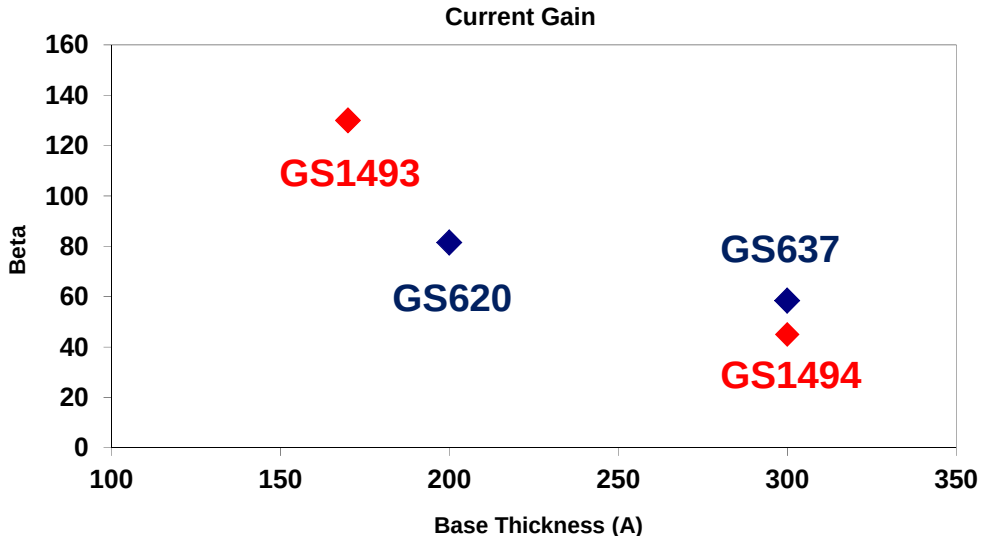

Figure 4.2. Current gain versus base layer thickness.

The GS1493 has current gain over 130, which is promising to show the potential of Type-I emitter-base alignment plus graded base structure. The hot electron injection with built-in electric field gives the carrier higher kinetic energy. With decent base sheet resistance and current gain nearly twice as high, the Type-I/II DHBTs have potential to reach higher f_T / f_{MAX} simultaneously. Because the emitter and subcollector designs are fixed, the properties of these layers are relatively constant across the structures. Typical emitter sheet resistance is 55 Ω/□ with specific contact resistance 2-5 Ω-μm^2. Typical subcollector sheet resistance is 2-3 Ω/□ with specific contact resistance 75 Ω-μm^2.

Figure 4.3 shows the large-area forward Gummel I-V curve for GS1493 and GS620. Figure 4.4 shows the forward Gummel plots for GS1494 and GS637. The large area-to-perimeter ratios of the 50×50 μm^2 devices ensure that intrinsic space-charge region and bulk base recombination effects dominate the behavior rather than extrinsic surface or contact recombination. In the forward-active mode, the base current ideality gives insight into the amount of space-charge versus bulk recombination, while the collector current contains information about electron injection into and transport through the base layer [30].

Looking first at the forward Gummel plot of the comparison of GS1493 and GS620, the base current of GS1493 is larger at low bias with a flatter slope (high ideality factor). This corresponds to space-charge recombination. However, the crossover point of I_C and I_B is 0.05 V lower than that of GS620. The space charge recombination in the emitter-base junction in GS1493 is more significant than that in GS620 because the strained InAlP growth is not optimized. But the reduced crossover point is a sign of localized electron pile-up in a Type-II alignment interface that has been broken in the raised conduction band edge in a Type-I emitter-base interface. The collector ideality factor of GS1493 is slightly larger than GS620, indicating

that the electron current is limited by the emitter-base conduction band barrier, which is expected in the Type-I junction. The emitter-base diode and base-collector diode curves are the other measurements to tell the junction characteristics of materials. Under reverse bias the base-collector diode shows 10^4 times higher current in GS1493 compared to GS620. This can be attributed to the higher collector doping concentration. Similar features between GS1494 and GS637 are observed in Figure 4.4.

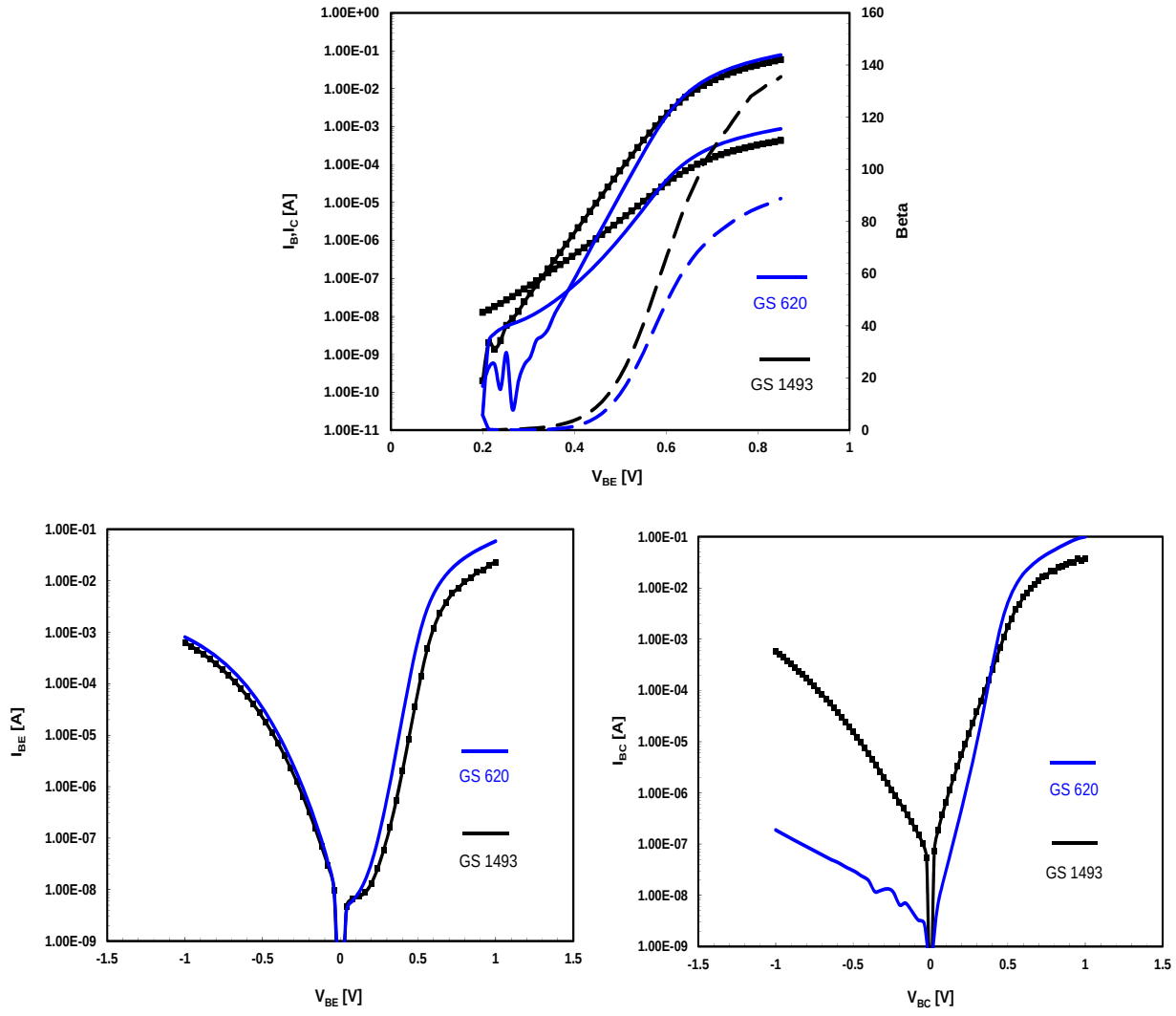

Figure 4.3. Forward Gummel I-V curves, EB, and BC diode for large area 50 × 50-μm^2 devices of GS1493 and GS620.

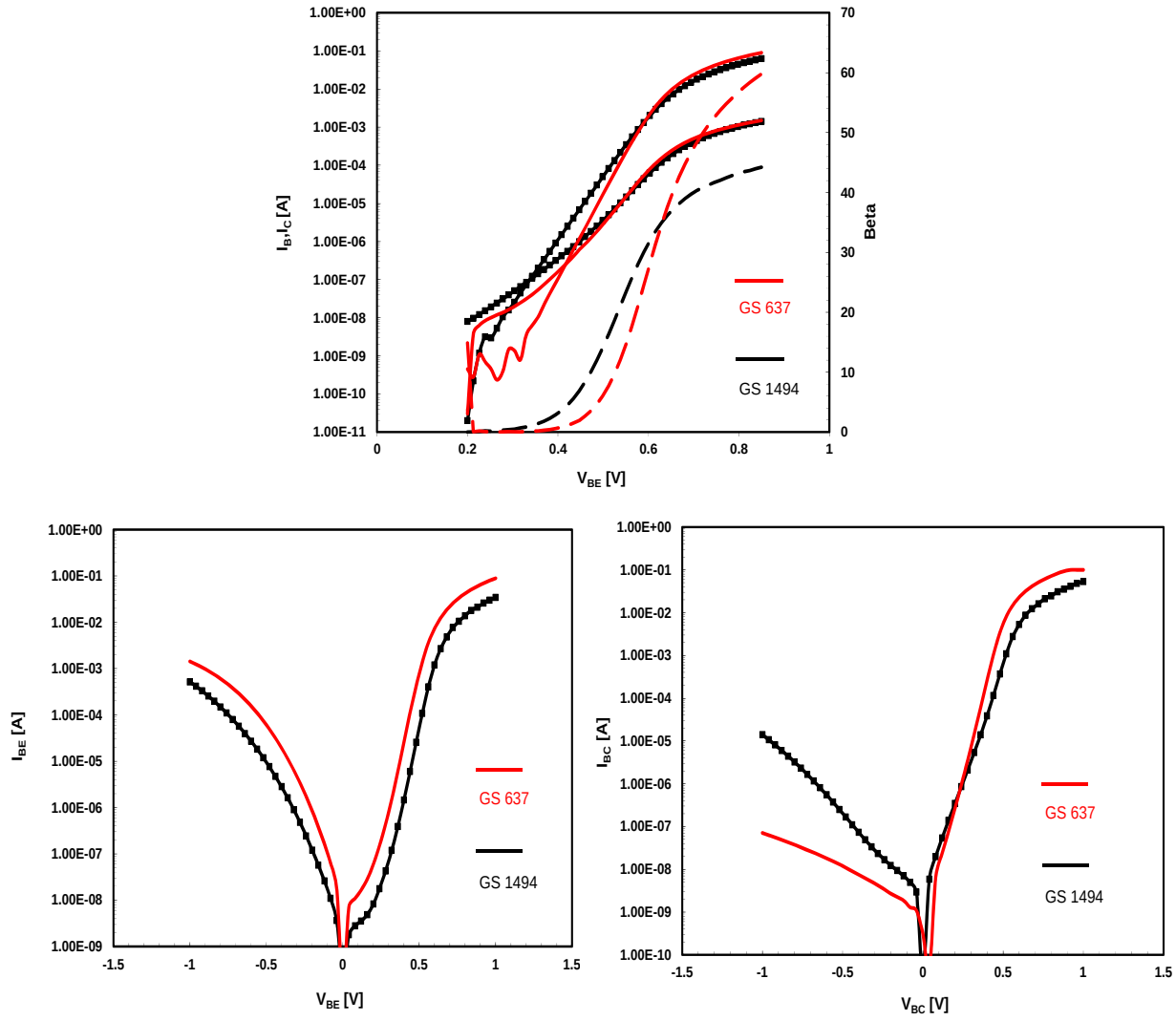

Figure 4.4. Forward Gummel I-V curves, EB, and BC diode for large area 50 × 50-μm^2 devices of GS1494 and GS637.

4.2 Yield Analysis of Submicron Devices

The initial results of submicron devices suffer from a low yield issue. About 10% of the devices can demonstrate normal HBT behavior. But repeatedly measuring the same device induced a destructive effect which was never observed in successful lots. The base-emitter and base-collector diode curves of several 0.5×5 μm^2 devices are shown in

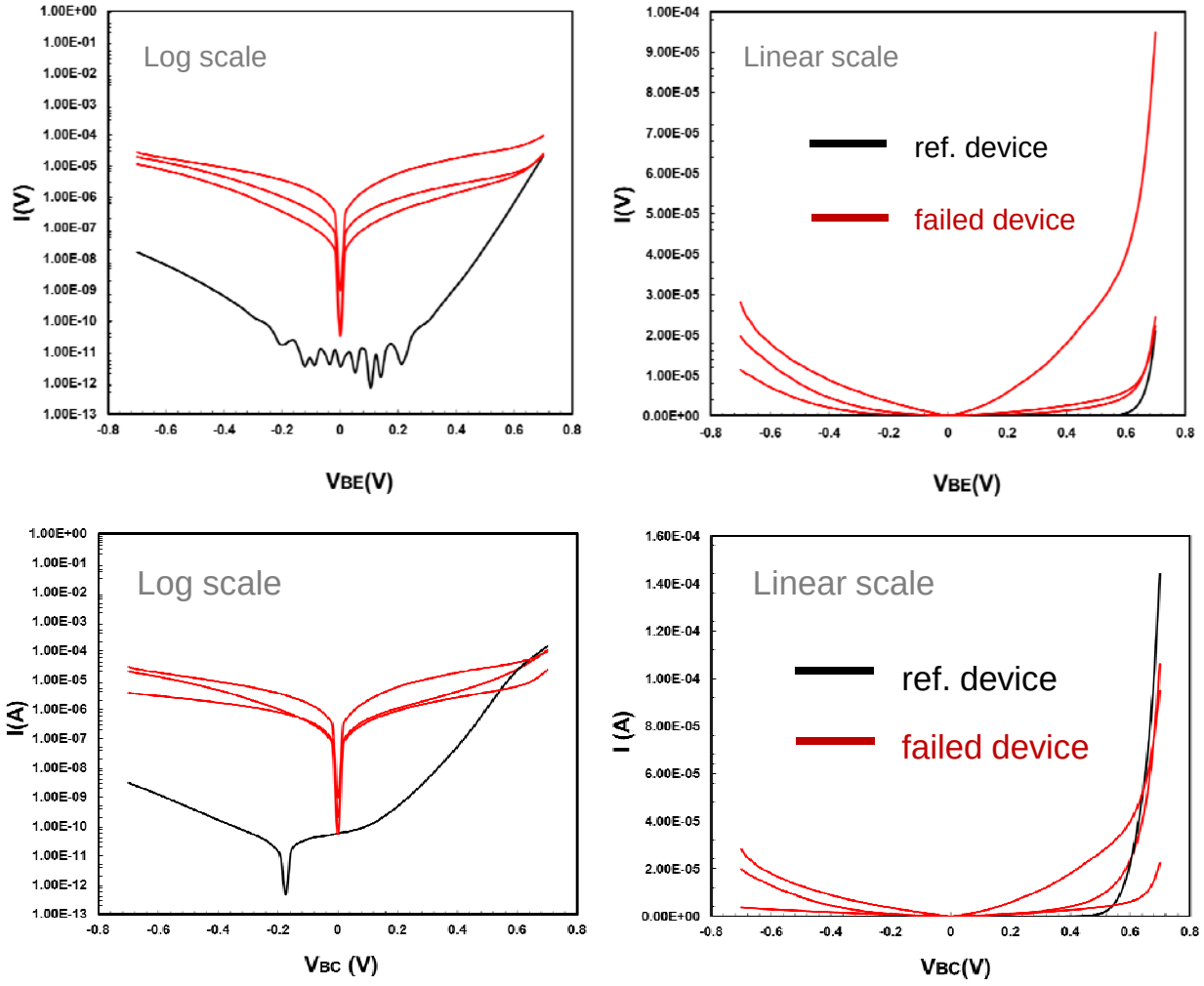

Figure 4.5. Diode curves of 0.5×5 μm^2 devices in the low yield lot (red) compared with reference devices (black).

Figure 4.5. Compared to the reference working device data, both diodes show leaky characteristics. The low yield is more likely due to global damage during the process. To identify which step creates the problem, a focused ion beam tool was used to check the cross section of the submicron device. The SEM picture of the device cross section shows the base layer has been attacked and etched all the way from the edge of the base metal into the intrinsic region. This damage explains the low current gain and high emitter contact resistance of working devices.

4.3 Submicron Device DC Data

Figure 4.6 [8] shows the collector current density (J_C-V_{CE}) characteristics of a 0.35 x 4 μm^2 emitter Type-I/II DHBT for I_B = 0 to 350 µA with a step of ΔI_B = 35 µA. The knee voltage is less than 0.65 V at J_C = 12 mA/μm^2, while the offset voltage is 0.1 V. The collector-emitter breakdown voltage of BV_{CEO} = 4.2 V is determined at J_C = 100 A/cm^2 with I_B = 0. The Gummel curve exhibits a peak current gain greater than 50, as is shown in the inset of Figure 4.6. The crossover point of I_B and I_C is at V_{BE} = 0.48 V.

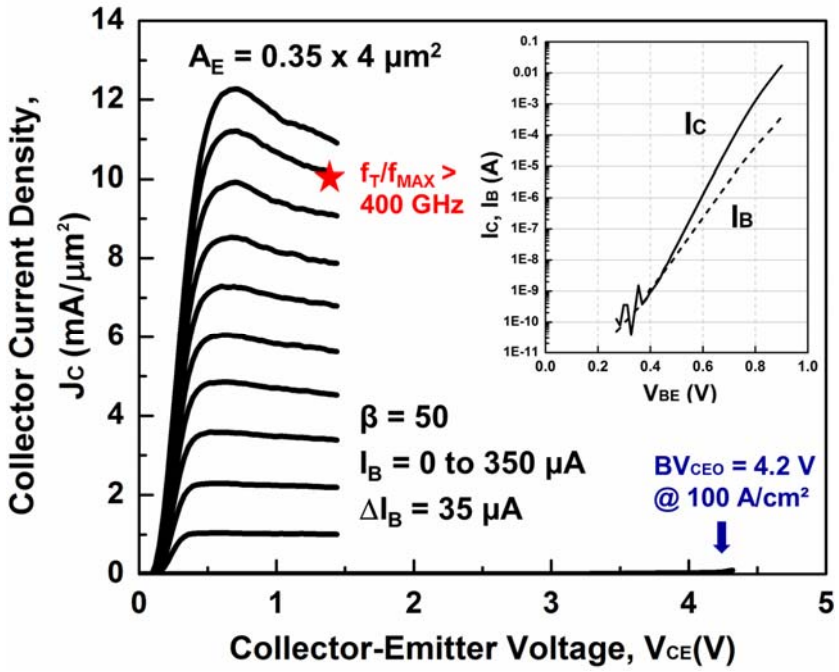

Figure 4.6. Emitter current density and Gummel (inset) of I-V characteristics for a 0.35 × 4 μm^2 Type-I/II Sb DHBT. Measurements are taken at room temperature, and indicate BV_{CEO} = 4.2 V. Peak speed (f_T and f_{MAX}) is measured at current density, J_C =10 mA/μm^2 and V_{CE}=1.35 V.

4.4 Submicron Device RF Data

Device microwave performance was measured from 0.5 GHz to 50 GHz using an Agilent E8364A network analyzer. Calibration was performed using off-wafer short-open-load-through

(SOLT) calibration, and on-wafer open and short standards are used to de-embed pad capacitances and inductances from the device S-parameters. Current gain, H_{21}, and Mason's U for a 0.35 x 4 μm^2 device are shown in Figure 4.7 [8]. The cutoff frequencies are extrapolated using a least-square fit of -20 dB/decade lines and a single pole approximation. The peak current gain cutoff frequency (f_T) is 455 GHz and the unilateral power gain cutoff frequency (f_{MAX}) is 400 GHz at a current density of J_C = 10 mA/μm^2 with V_{CB} = 0.5 V and V_{CE} = 1.35 V. Figure 4.7 (inset) shows the device cutoff frequency for devices with different emitter lengths at V_{CB} = 0 V. f_T increases for longer emitters and f_{MAX} shows the opposite trend. The longer emitter length provides a reduction of emitter contact resistance and results in the reduction of minority and junction charging times; hence, f_T increases with emitter length L_E. The reduction of emitter length equals the reduction of C_{BC} and results in an increase in f_{MAX}.

An equivalent small signal circuit model for devices presented in this work was extracted from a combination of S-parameter measurements and device geometry [31]. Figure. 4.8 [8] illustrates the T-model of the small signal equivalent circuit and gives the component values extracted for the 0.35 x 4 μm^2 device with f_T = 520 GHz. The simulated scattering parameters agree well with the measured 0.5-50 GHz device data. Equation (4.1) identifies the components of f_T delay for an HBT. Here τ_B and τ_C are the base and collector transit times. R_{EE}, R_B, and R_C are the emitter, base and collector resistances, respectively. C_{JE} is the base-emitter junction and emitter diffusion capacitance and C_{BC} is the base-collector junction capacitance. The $r_E = \eta kT / qI_C$ term is the dynamic resistance of the forward-biased emitter-base junction.

$$\frac{1}{2\pi f_T} = \tau_B + \tau_C + \frac{\eta kT}{qI_C} C_{JE} + (R_C + R_{EE} + \frac{\eta kT}{qI_C}) C_{BC} \tag{4.1}$$

Plotting $\frac{1}{2\pi f_T}$ versus $\frac{1}{I_C}$, extrapolating to $\frac{1}{I_C} = 0$, and subtracting $(R_{EE} + R_C) \times C_{BC}$ yields the base-collector transit delay, $\tau_B + \tau_C$. The slope of this plot is used to determine C_{JE}. Once these parameters are known, the total device delay can be split into the collector charging time $\tau_{CC} = (r_E + R_{EE} + R_C) \times C_{BC}$, the emitter charging time $\tau_E = r_E \times C_{JE}$, and the base-collector transit delay, $\tau_B + \tau_C$. The total device delay time τ_T is 309 fs. The base and collector transit time $\tau_B + \tau_C$ is 196 fs. The emitter charging time τ_E is 25 fs and collector charging delay τ_{CC} is 88 fs. The base transit time and collector transit time can be further partitioned. The Type-I/II DHBTs with a 17 nm ternary GaAsSb base in this work have base transit time τ_B of 67 fs while the 25 nm quaternary InGaAsSb graded base DHBT showed τ_B = 122 fs [6]. The effective base velocity of the Type-I/II DHBT is 2.54×10^7 cm/sec compared to 2.05×10^7 cm/sec in InGaAsSb Type-II DHBTs. Despite the composition graded base of the ternary Type-I/II configuration ($\Delta E_{C,ternary}$ = 20 meV) being less than the quaternary Type-II configuration ($\Delta E_{C,\ quaternary}$ = 50 meV), the Type-I/II DHBT sees a higher effective velocity as a result of the hot electron injection of the Type-I band alignment at base-emitter junction with $\Delta E_{C,\ EB\ junction}$ = 140 meV. This hot electron injection provides the fast base transit time, τ_B = 67 fs, and improved current gain, β = 50. Furthermore, carriers injected from Type-II base/collector junction exploit velocity overshoot in the collector layer. The average effective collector velocity is 3.76×10^7 cm/sec from electron transport modeling, which is higher than bulk saturation velocity in InP (2.6×10^7 cm/sec).

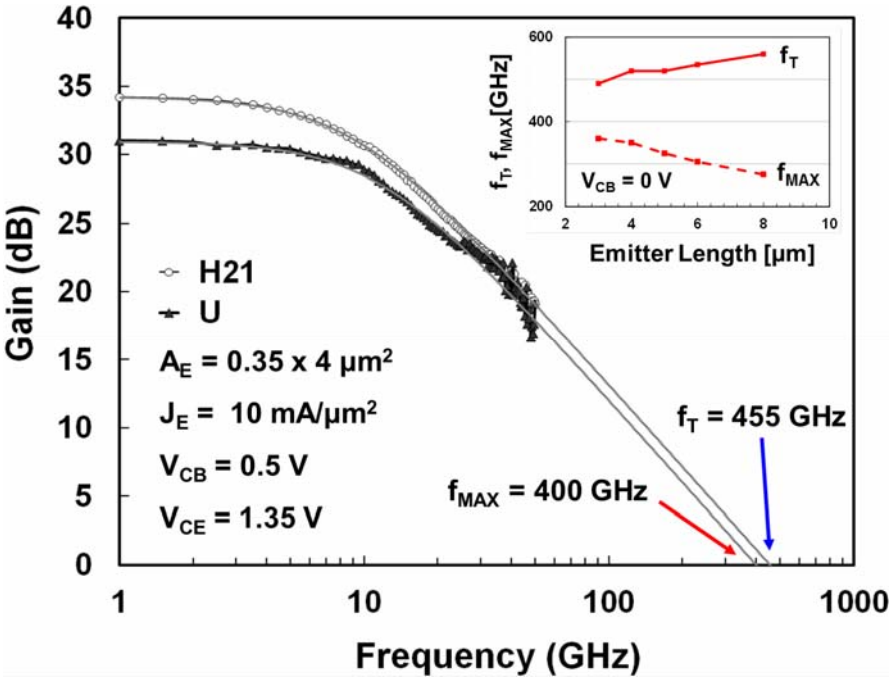

Figure 4.7. Current gain H21 , and Mason's unilateral gain U extrapolations. f_T = 455 GHz and f_{MAX} = 400 GHz for 0.35 × 4 μm² DHBT at V_{CB}= 0.5 V using -20 dB/decade method. The inset shows the plot of f_T (solid) and f_{MAX} (dash) versus emitter width at V_{CB}= 0 V.

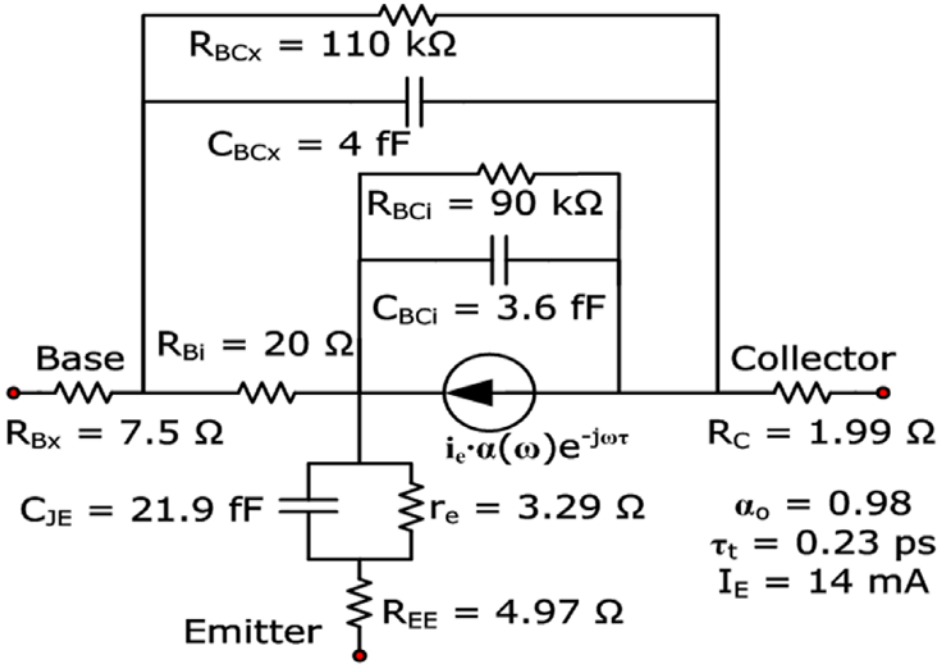

Figure 4.8. Small-signal equivalent circuit model for a 0.35 × 4 μm2 Type -I/II Sb DHBT with peak f_T = 520 GHz based on the measured and fitted S-parameters from 0.5 to 50 GHz.

5 BENCHMARK OF TYPE-I/II DHBT WITH TYPE-I DHBT

5.1 Current Blocking in Type-I DHBT

InP-based HBTs possess higher electron saturation velocity, higher breakdown electric field if using InP collector, lower emitter-base junction turn-on voltage (better power added efficiency), and higher themal conductance of InP substrate. These advantages over GaAs HBTs make them promising to dominate the next generation of communication systems. However, InP SHBTs suffer from low breakdown voltage. Mixed signal integrated circuits require transistors exhibiting high linearity and high breakdown voltage to improve dynamic range for practical applications. InP double heterojunction bipolar transistors, therefore, have drawn attention because incorporating wide-gap InP as the collector layer offers the potential to improve the breakdown behavior. Unfortunately, in a typical Type-I band alignment DHBT there is a conduction band discontinuity between the InGaAs base and InP collector layer. This energy barrier leads to the so-called current-blocking effect in the collector I-V characteristics as well as lower current gain, which had been observed in Type-I InGaP/GaAs DHBTs. To mitigate the current-blocking issue, graded energy gap or chirped superlattice composite collectors were developed in Type-I DHBTs but with limited success at high current injection.

Due to the presence of conduction band edge discontinuity at the base-collector heterointerface, Type-I DHBTs have suffered from the current blocking issue. Various doping or compositional grading schemes therefore have been developed to eliminate the base-collector band alignment effect; these include grading the conduction band with an InAlGaAs quaternary alloy, or step grading from InGaAs to InP by inserting a InGaAs spacer. The foundry-provided Type-I DHBT devices we studied in this work utilized the concept of chirped superlattice structure to overcome the base-collector band edge discontinuity problem. The equilibrium

energy band diagram of a Type-I DHBT is shown in Figure 5.1: the transition region is composed of a setback layer, a chirped superlattice, and a delta-doping layer. The p-type setback layer is intended to reduce the energy barrier via using narrow bandgap InGaAs at the expense of base pushout at high current density and lowered breakdown voltage. The InAlAs/InGaAs chirped superlattice with a variable duty cycle is used to engineer minibands that create a linear graded region. The resultant average composition changes linearly from InGaAs to InAlAs. The pulse doping (Si: 5×10^{18} cm^{-3}) near the InP collector together with the heavily doped base layer provides necessary sheet charge to form the dipole. The dipole-induced electrostatic field is supposed to cancel the field created by the linear bandgap grading in chirped superlattice layers. In the next sections, we will show that, even with all the complicated epigrowth schemes described above, Type-I DHBT with composite collector still fails to minimize the trapping of carriers in the conduction band notches presented at the base-collector junction.

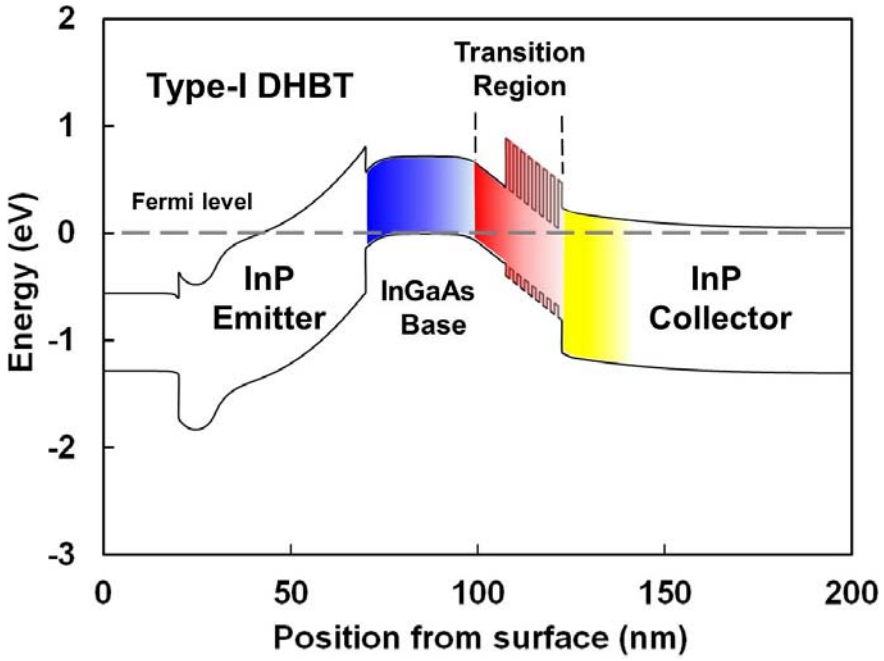

Figure 5.1. Energy band diagram for a Type-I DHBT with a composite collector. A setback layer and superlattice transition region are inserted between the base and InP collector layer.

5.2 Knee Voltage, Gain Compression and Breakdown Voltage

As the transistor is used as an amplifier, smaller knee voltage progression is desired to achieve linear amplification at lower bias voltage. It is also important to maintain constant current gain (β) in operation range for maximum allowable signal swing at the collector. Figure 5.2 shows the I-V collector current density (J_C-V_{CE}) characteristics of a Type-I/II GaAsSb DHBT (red) and Type-I InGaAs DHBT (black). The Type-I/II device demonstrates lower knee voltage (less than 0.62 V at J_C = 10 mA/μm^2) than the Type-I device with the same biased current. At high current density, the Type-I InGaAs DHBT exhibits considerable current gain compression, unlike the Type-I/II GaAsSb DHBT, which also has higher β. The devices have two emitter sizes: A_E = 0.35×3 μm^2 and A_E = 0.5×3 μm^2. The slightly negative slope of the Type-I/II device is caused by the heat generated at high biased current density (J_C > 5 mA/μm²).

Figure 5.3 illustrates the gain compression shown in the above family curves: the Type-I/II devices show less than 5% reduction or nearly constant β, while Type-I devices suffer 20% β variation in the forward-active region (V_{CE} = 1 V). Either high knee voltage or gain compression is related to nonlinear DC behavior. The main cause of these phenomena is the BC junction alignment difference between Type-I and Type-I/II devices. As Figure 5.1 specifies, despite the proposed concept of composite collector to overcome current blocking in Type-I DHBTs, the transition region is not able to eliminate the carrier transport impedance, which leads to nonideal DC performance.

Conventionally, breakdown voltage needs to be two to four times higher than supplied voltage for practical circuit applications. However, in order to enhance speed by vertically scaling material, the breakdown voltage of the device is lowered dramatically. InP DHBTs, among all InP-based transistors, show potential to solve the speed-breakdown tradeoff. In this

work, a Type-I/II device with T_C = 1000 Å exhibits BV_{CEO} = 3.8 V determined at J_C = 1 kA/cm^2 with I_B = 0 (Figure 5.4). The Type-I device with total T_C = 1500 Å (transition region: 250 Å, InP 1250 Å), on the other hand, shows slightly higher BV_{CEO} = 3.92 V (Figure 5.5). For fair comparison, Type-I/II DHBTs demonstrate higher collector-emitter breakdown voltage *per unit of collector layer thickness* due to the all-InP collector design. The composite collector of Type-I DHBTs lowers the breakdown because the narrow bandgap material is used as the setback layer and superlattice structures.

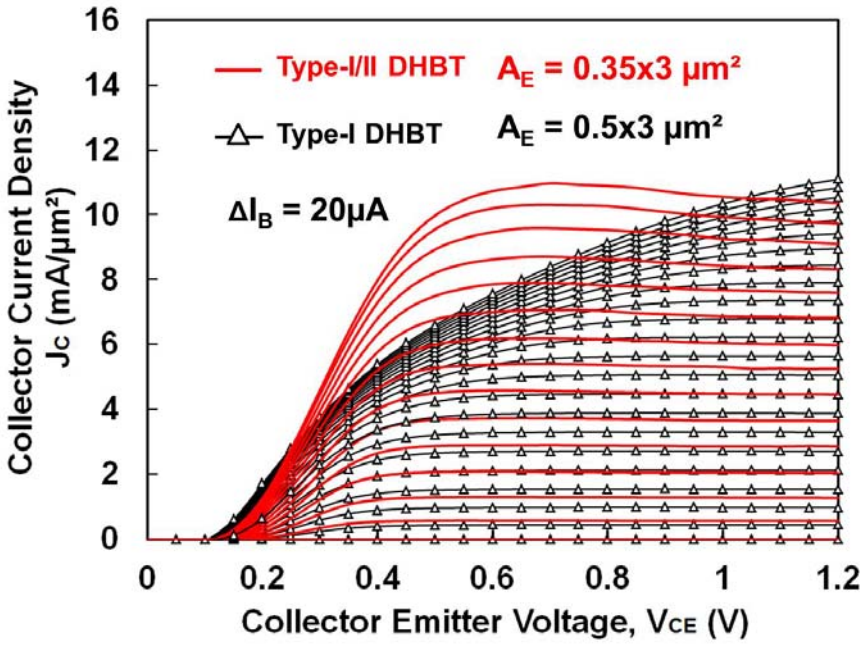

Figure 5.2. Collector I-V of UIUC Type-I/II DHBT (red) and foundry Type-I (black) InP DHBT. The foundry device exhibits considerable β compression due to current blocking at B/C junction despite the composite collector transition layer.

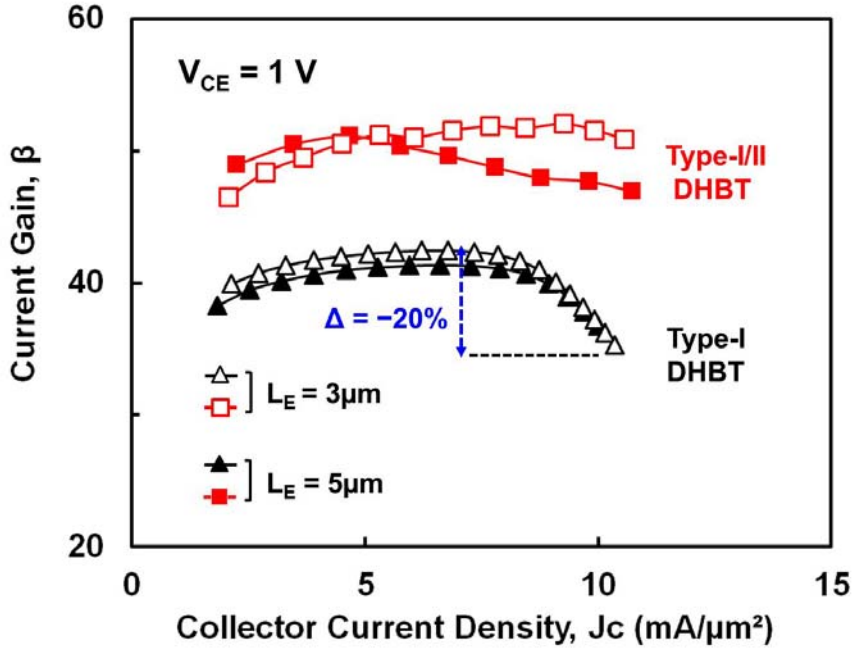

Figure 5.3. Type-I DHBT exhibits β compression with >20% drop in current gain at high current density. UIUC Type-I/II DHBT shows minimal β variation. Current gain compression is related to nonlinearity in device operation.

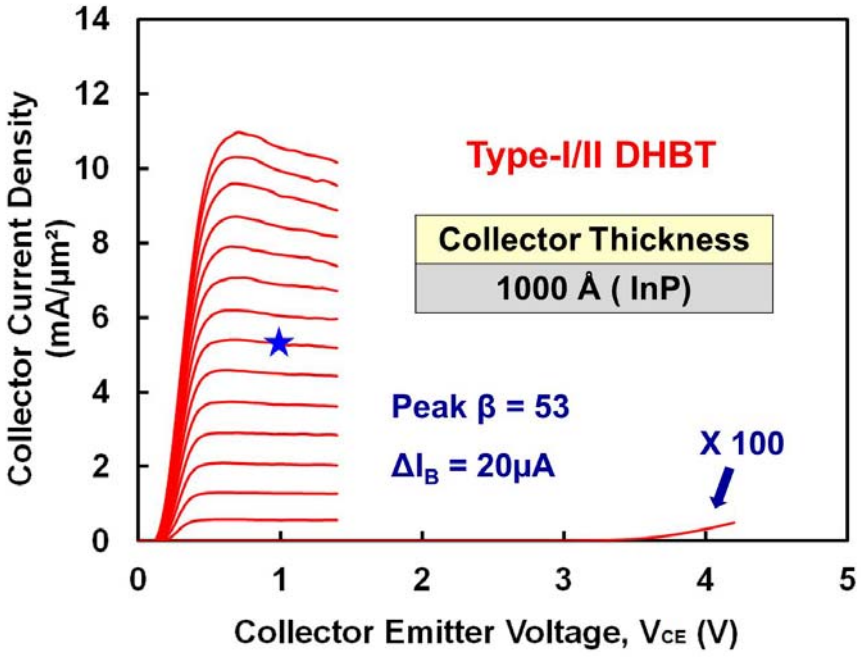

Figure 5.4. Common-emitter family curves of Type-I/II DHBTs show gain = 53 and BV_{CEO} = 3.8 V with all-InP collector thickness of 1000 Å.

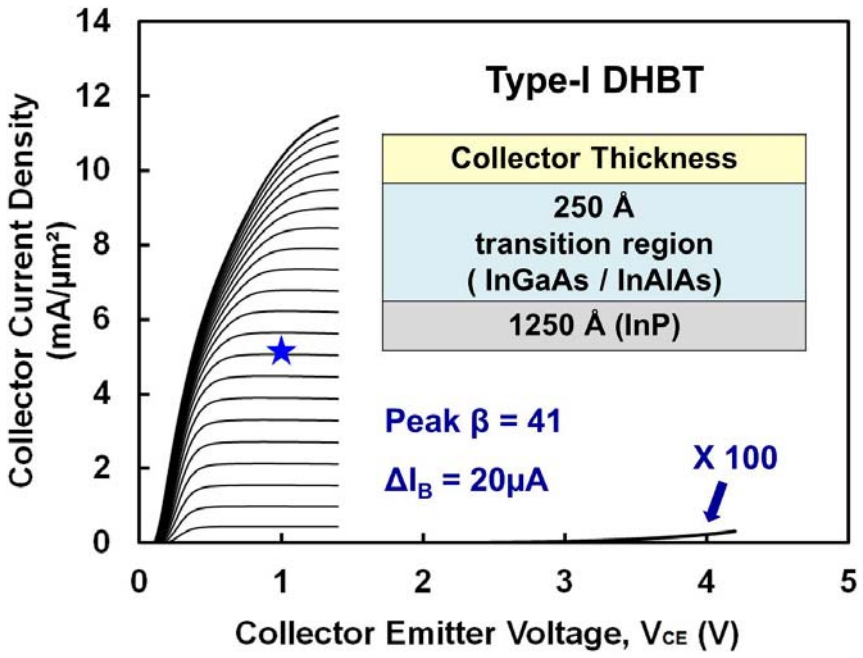

Figure 5.5. Common-emitter family curves of Type-I/II DHBTs show gain = 41 and BV_{CEO} = 3.92 V with composite collector thickness of 1500 Å.

5.3 High Current Injection Effect and Speed Falloff

Figure 5.6 shows the current gain cutoff frequency (f_T) versus collector current density (J_C) of Type-I/II DHBT (red) with device area $A_E = 0.35\times3$ μm^2 and Type-I DHBT (blue) with $A_E = 0.5\times3$ μm^2. The Type-I DHBT demonstrates speed falloff with increasing J_C, while the f_T of Type-I/II DHBT keeps increasing with J_C until saturated as J_C > 10 mA/μm². And similar speed falloff of the power gain cutoff frequency (f_{MAX}) is presented in Figure 5.7.

At high current densities, the electron concentration in the collector layer can rise to a level comparable to that of the collector doping. The positive charge of the uncovered donors is thus compensated by the electron charge, and the electrostatic field in the depleted collector layer will begin to decrease. At the end, the mobile electron concentration can exceed the collector doping concentration and the electric field will drop to zero near the base-collector junction. In a homojunction case, the holes from the base will move into the zero field portion of the collector,

extending the base width. This base push-out, will increase the base transit delay, and lower the current gain.

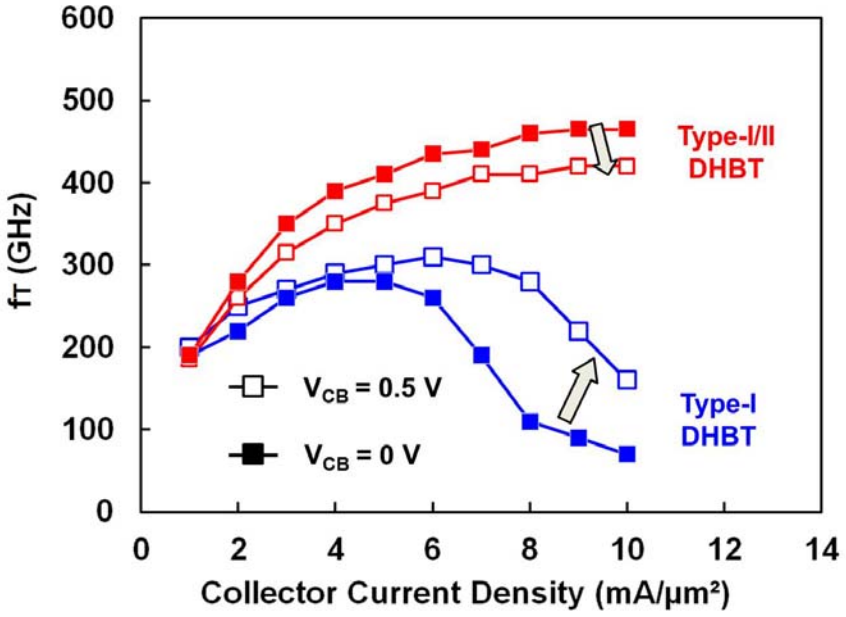

Figure 5.6. Type-I devices show f_T rolloff behavior at $J_C > 6$ mA/μm^2. The carrier transit time in the base is increased due to current blocking. The nonlinearity is related to 2nd derivative of f_T with respect to collector current density (J_C).

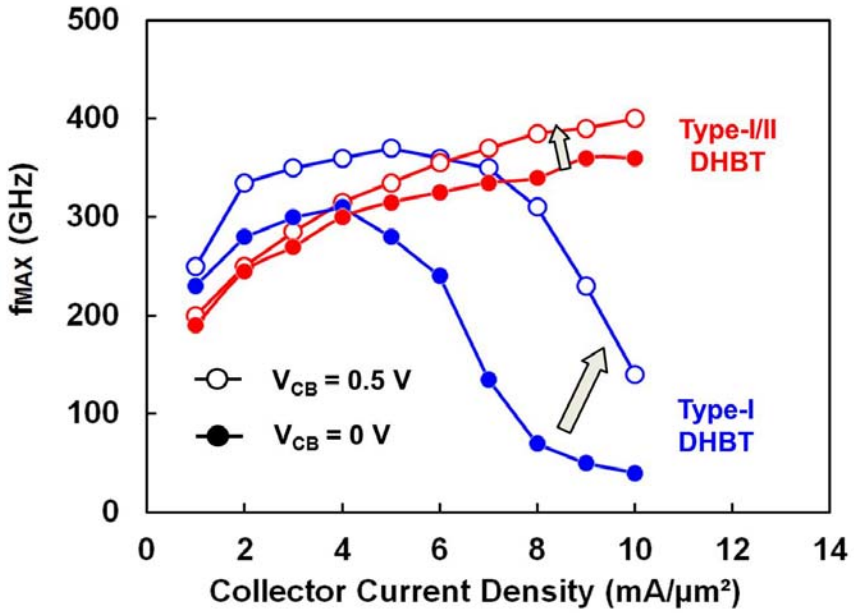

Figure 5.7. Power gain cutoff frequency f_{MAX} vs. collector current density of devices with A_E = 0.35×3 μm^2 at V_{CB} = 0 V and 0.5 V. Foundry Type-I devices show rolloff with increasing J_C and exhibit power gain nonlinearity.

However, in the case of DHBTs which have a base-collector heterojunction, the built-in barrier will prevent base push-out. This means that the Kirk effect cannot sufficiently explain the speed-J_C relation observed in DHBTs studied in this work. Using a two-dimensional hydrodynamic numerical simulator enables us to effectively investigate the RF performance differences between the two types of DHBTs at high current injection. We exploited the Sentaurus Device (Synopsys, Inc.) package which is able to self-consistently solve both the Poisson and transport equations to simulate the energy band diagram. Figure 5.8 illustrates the band diagram of Type-I DHBTs at V_{CB} = 0 V and 0.5 V, respectively. When injecting current density rising from 4 mA/μm^2 to 10 mA/μm^2, both conduction and valence band edges are pulled up. The energy barrier at the composite collector layer is therefore raised as well, trapping more minority carriers in the neutral base region. As Figure 5.8 shows, the carriers are still impeded and bounced back and forth although the superlattice structure was intended to assist electron tunneling through the InAlAs layer. Secondly, the 15 nm thick p-type InGaAs setback layer, like the base material, makes the base-collector junction essentially a homointerface. The neutral base spreading occurs accordingly. When J_C = 6 mA/μm^2, the f_T reaches maximum value and starts dropping after that. The base push-out and superlattice barrier lead to minority charge accumulation in the base, as shown in the carrier distribution plot in Figure 5.8. Type-I/II DHBT, in contrast, shows speed increases with J_C. The band diagram in Figure 5.9 clearly demonstrates that the all-InP collector layer enables carriers to travel smoothly through the BC junction. As J_C attains 10 mA/μm^2, the band edge is raised and flattened in the collector depletion region near the BC interface. Type-II BC alignment ($\Delta E_{C,\,Base} > \Delta E_{C,\,collector}$) counters the energy band up-stretch, alleviating the impact of high current injection. However, as J_C is

high enough to raise the conduction band edge in the collector over that of base, speed falloff will still happen in Type-I/II DHBT.

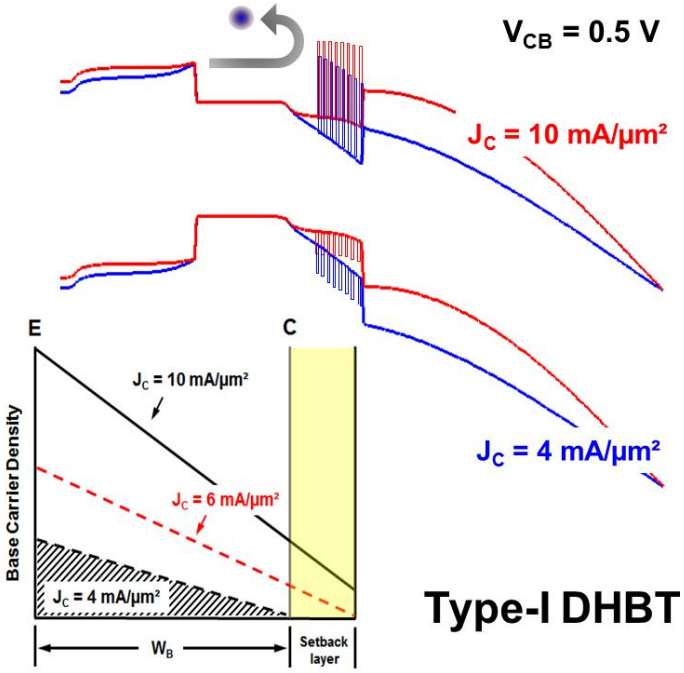

Figure 5.8. Band diagram and carrier distribution of Type-I DHBT at different J_C.

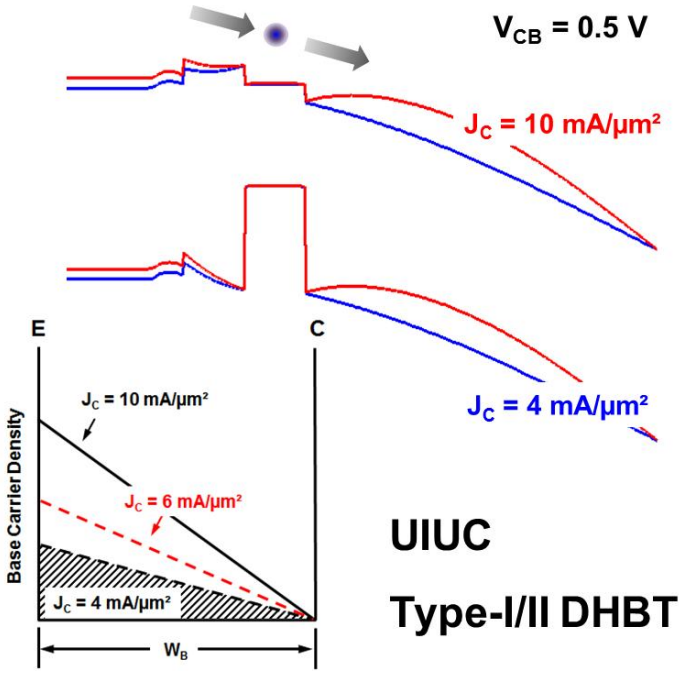

Figure 5.9. Band diagram and carrier distribution of Type-I/II DHBT at different J_C.

When varying V_{CB} from 0 V to 0.5 V, f_T of the Type-I DHBT improves because reverse biasing the base-collector junction effectively pulls down the energy barrier, allowing electrons to be collected without blocking. Type-I/II DHBT, however, shows f_T reduced slightly while V_{CB} increases. Since the base-collector junction favors electron transport, the applied bias accelerates electrons in the depleted collector layer with higher field strength. As electrons acquire sufficient energy to transfer from the high mobility Γ-valley to the low mobility L-valley, the effective velocity in the collector v_{eff} is reduced. It is calculated that $v_{eff} = 3.06\times10^7$ cm/sec at $V_{CB} = 0$ V and $v_{eff} = 2.35\times10^7$ cm/sec at $V_{CB} = 0.5$ V. On the other hand, f_{MAX} of both DHBTs is enhanced with higher BC bias voltage. The relation between f_T and f_{MAX} can be expressed as $f_{MAX} = \sqrt{f_T / 8\pi \times C_{BC} \times R_B}$. C_{BC} is the base-collector junction capacitance and R_B is the base-sheet resistance. The collector layer depletion region is widened with V_{CB}, contributing to lower C_{BC} and higher f_{MAX}.

5.4 Delay Time Analysis and Microwave Parameter Extraction

From the same procedure to extract small-signal parameters described in section 4.4, four time constants constituting the total delay time can be determined. Figure 5.10 and Figure 5.11 depict four delay time constants versus J_C of Type-I DHBT ($A_E = 0.5\times3$ μm^2) and Type-I/II DHBT ($A_E = 0.35\times3$ μm^2), respectively. τ_B and τ_C of Type-I DHBT are longer than those of Type-I/II DHBT owing to thicker epilayer which causes longer transit time. The base transit time is further modified by current gain variation if we assume the base current is mainly composed of recombination current in the neutral base region. And the ratio of collector current to base current is $I_C / I_B = \beta = \tau_n / \tau_B$, where β is equal to the ratio of electron carrier lifetime in base (τ_n) to base layer transit time (τ_B). Type-I DHBT obviously suffers more from

beta compression at higher J_C, leading to increased τ_B compared to nearly constant τ_B with J_C in Type-I/II DHBT.

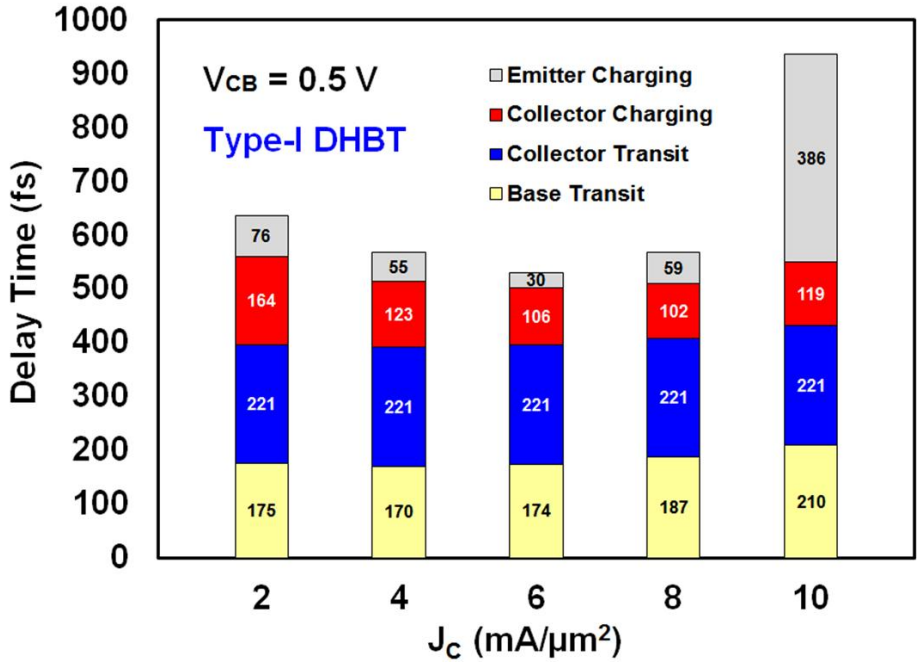

Figure 5.10. Delay time constants vs. J_C of Type-I DHBT.

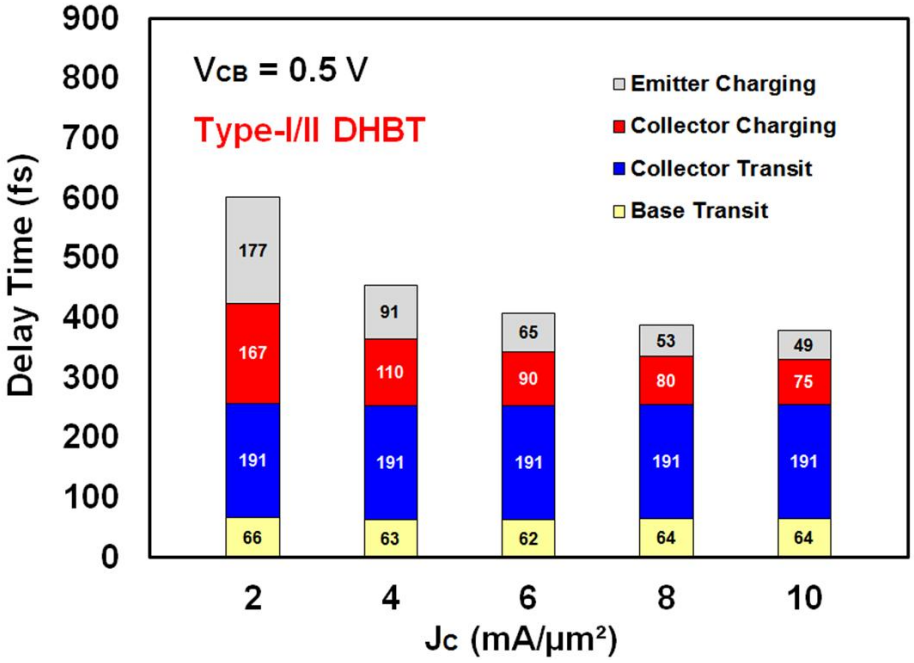

Figure 5.11. Delay time constants vs. J_C of Type-I/II DHBT.

It is apparent that what distinguishes the delay time deviation between Type-I and Type-I/II DHBTs is charging time constants. The emitter charging time τ_E fluctuation accounts for the delay time change with J_C. Figure 5.12 shows C_{JE} and r_E (inset) from J_C = 1 mA/μm^2 to J_C = 10 mA/μm^2. r_E drops with increasing J_C for both devices (operating at high current reduces charging delay, offsetting the effect of vertically scaled HBT). Type-I/II DHBT has larger r_E because of smaller contact area than Type-I DHBT. At J_C < 7 mA/μm^2, C_{JE} keeps almost constant for both types of DHBT. But C_{JE} of Type-I DHBT surges exponentially to over 200 fF, as in Type-I/II DHBT C_{JE} increases gradually from 9 fF to 15 fF. The diffusion capacitance $C_{JE\text{-}diff} = \partial Q_{diff}/\partial V_{BE}$, where Q_{diff} is the minority carriers (electrons here) caused by forward-biasing BE junction. The InAlAs/InGaAs superlattice and setback layer of the composite collector in Type-I/II DHBT trap the charge inside the base and transition region. With high current injection, the excessive Q_{diff} contributes to soaring $C_{JE\text{-}diff}$, and total emitter-base junction capacitance C_{JE}.

The change of collector charging time τ_{CC} versus J_C is mainly due to the base-collector junction capacitance C_{BC}. Figure 5.13 shows that C_{BC} of Type-I/II DHBT decreases progressively, but Type-I DHBT shows C_{BC} is getting higher with J_C. From $\tau_{CC} = (r_E + R_{EE} + R_C) \times C_{BC}$, the τ_{CC} decreases in Type-I/II due to the smaller emitter resistance and base-collector capacitance simultaneously, while in Type-I DHBT, C_{BC} makes collector charging time longer at high J_C. It is also clear at V_{CB} = 0 V that C_{BC} is almost three times larger than at V_{CB} = 0.5 V because reverse bias voltage widens depletion region width and lowers the BC alignment barrier height that reduces charge trapping. The C_{BC} variation at different V_{CB} also explains why Type-I DHBT shows f_{MAX} enhancement more than Type-I/II DHBT as V_{CB} attains 0.5 V. Both C_{JE} and C_{BC} are related to charge storage effect. The composite collector

design of Type-I DHBT induces carrier trapping in the base region. Both junction capacitances exhibit nonlinearity in terms of operating current density. The charge storage effect has not only great influence on DC and RF performance but also on linearity characteristics.

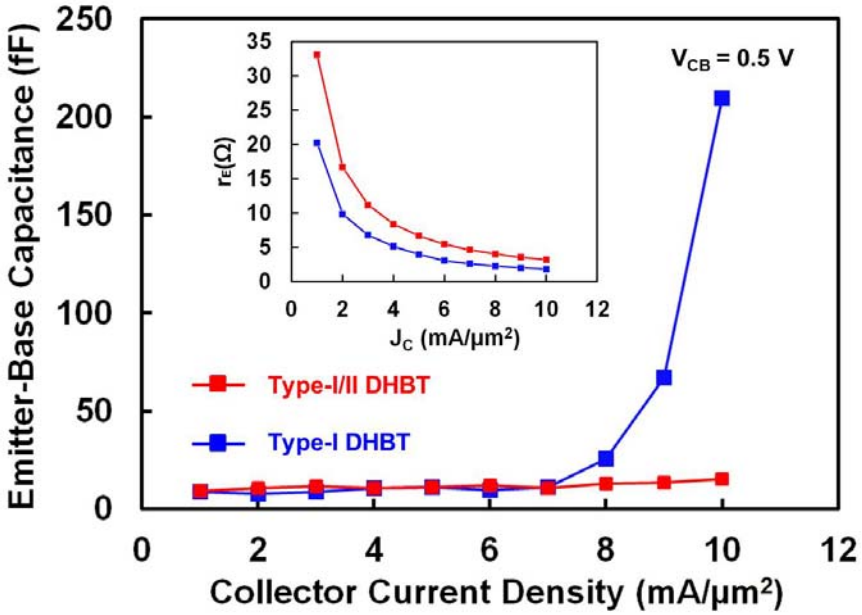

Figure 5.12. Extracted emitter-base junction capacitance and (inset) emitter dynamic resistance (r_E) vs. collector current density of Type-I and Type-I/II DHBT at V_{CB} = 0.5 V.

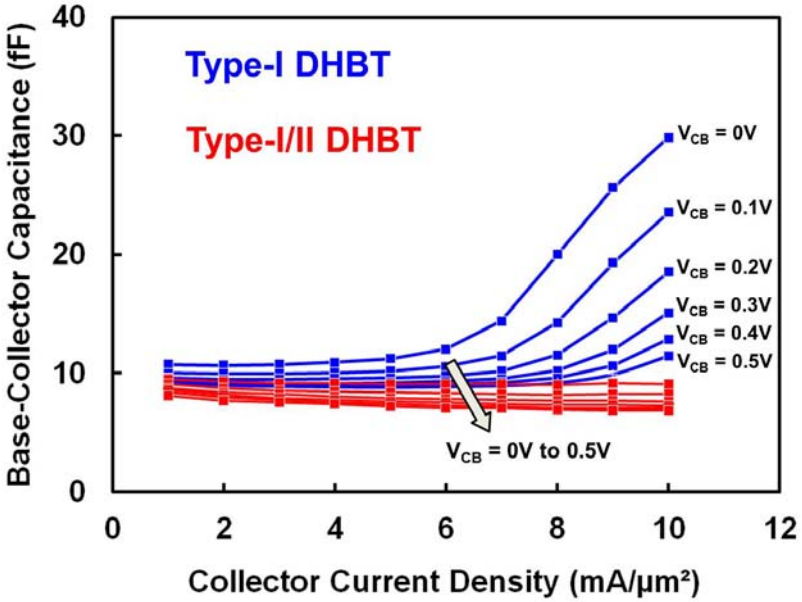

Figure 5.13. Extracted base-collector junction capacitance vs. collector current density of Type-I and Type-I/II DHBT at V_{CB} = 0 V to 0.5 V.

5.5 Microwave Linearity due to the Influence of Base-Collector Alignment

For an appropriate comparison of device linearity, the operating gain (G_{OP}) compression was measured for each device under a constant current density and voltage bias. The operating gain is defined as the ratio of the power delivered to the load to the power input into the device, $P_{in} = P_{avs} \cdot (1 - |\Gamma_{in}|^2)$, where Γ_{in} is the input reflection coefficient of the device, as measured by the PNA. For this benchmarking measurement, the load was chosen to give the highest possible gain under stable operation for each device. Impedance tuning at the load was implemented with a 2-18 GHz Focus Microwaves mechanical tuner. The load impedances were chosen for each bias point using measured device scattering parameters to design for maximum gain inside the stable load impedance region. Power compression was measured at 12 bias points to view trends as a function of collector current density and V_{CE}. During measurement, source power (P_{avs}) was swept at 18 GHz using a power-calibrated Agilent E8364A PNA. Output power was measured using an HP 4419 power meter at the load side of the impedance tuner. The gain curves shown in Figure 5.14 are normalized for comparison of their 1-dB gain compression points. The devices were biased at constant collector current density and collector-emitter voltage, and the power sweep was performed at 18 GHz. Type-I/II DHBT has 7.75 dB linear power gain advantage over Type-I DHBT. Figure 5.15 shows the 1-dB power compression point ($P_{in\text{-}1dB}$) as a function of collector current density. $P_{in\text{-}1dB}$ is the input power for which 1-dB power gain compression is observed. At low current density, the two DHBTs have comparable linearity. But as current density increases, the Type-I/II devices show improved linearity. The Type-I device, however, shows a reduction in linearity as current density increases. This is attributed to the change in the emitter-base junction diffusion capacitance. At $J_C > 6$ mA/μm^2, the Type-I device linearity is greatly improved by increasing the voltage from 1.1 V to 1.4 V, more so than

at lower current densities. This again points to the base charge accumulation and base pushout effects, which are mitigated by increasing V_{CB} to pull down the Type-I BC junction energy barrier.

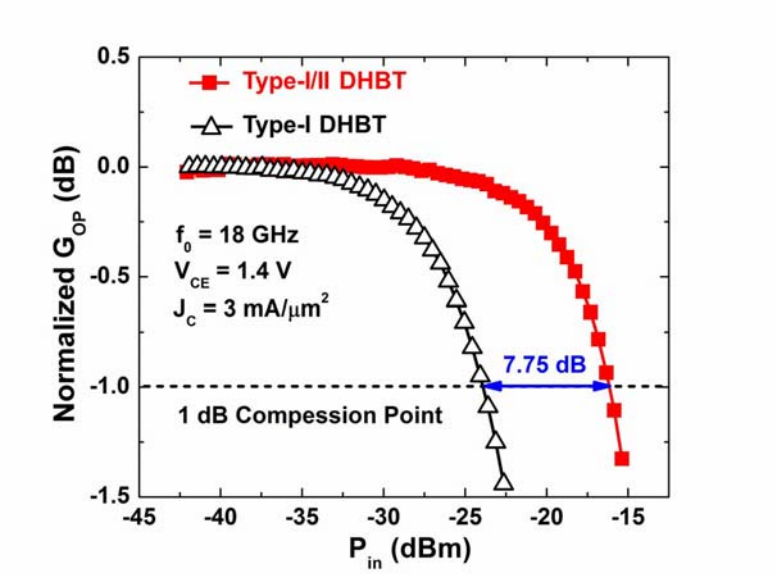

Figure 5.14. Operating gain versus input power, normalized to illustrate the 1 dB compression point for two devices.

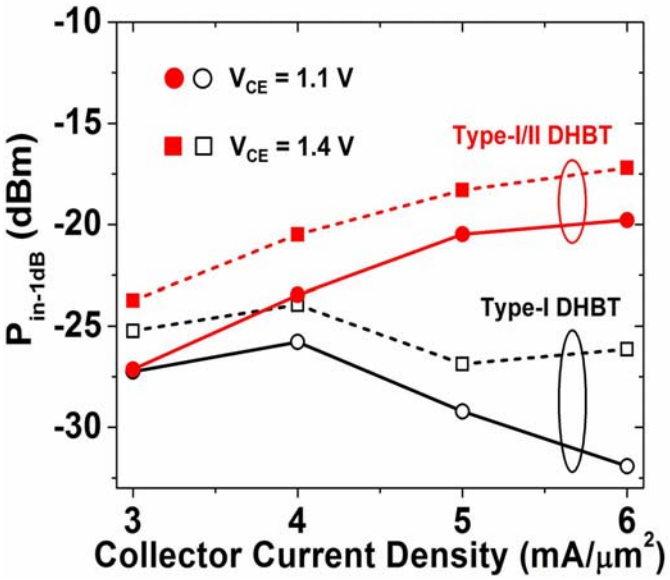

Figure 5.15. 1-dB power compression point (P_{in-1dB}) as a function of collector current density at different collector-emitter voltage (V_{CE}).

6 SUMMARY

We have developed hot electron injection submicron graded base Type-I/II Sb DHBTs and demonstrated base and collector velocity enhancement. The benefit of an abrupt emitter to create an energy launcher is clearly shown through the reduction in base transit time to enhance gain and frequency performance. These results strongly suggest that the graded base Type-I/II Sb-based DHBT device will play an important role for high speed mixed signal and millimeter wave integrated circuits. The persistent emitter wing issue was solved by optimization of both lithography and electron beam evaporator. The low yield problem was identified through various characterization tools and fixed in the sub-μm transistor fabrication. In addition, we have presented DC and RF measurement data on UIUC Type-I/II DHBT and foundry Type-I DHBT, showing the effect of the charge accumulation in the base and base pushout on the device speed performance. We demonstrate that the composite collector layer in Type-I DHBT creates severe carrier accumulation in the base. In terms of DC measurement, the gain compression is observed at high current injection for Type-I DHBT whereas Type-I/II shows minimum beta change. Due to the use of complicated transition layers, the breakdown voltage per collector thickness is lower in Type-I than Type-I/II DHBT. Microwave data illustrate the speed falloff of Type-I DHBT with increasing collector current density. Using small signal model parameter extraction and delay time constants decomposition analysis, the emitter diffusion capacitance, base transit delay and base-collector junction capacitance are attributed to transistor speed delay and microwave nonlinearity. Further linearity measurement demonstrates that high frequency distortion behavior is strongly related to the energy band structure design of Type-I and Type-I/II DHBTs.

REFERENCES

[1] J. Bardeen and W. H. Brattain, "The transistor, a semiconductor triode," *Physics Review*, vol. 74, pp. 230-231, 1948.

[2] W. Shockley, "Theory of p-n junctions in semiconductors and p-n junction transistors," *Bell System Technology Journal*, vol. 28, pp. 435-489, 1949.

[3] H. Kroemer, "Theory of a wide-gap emitter for transistors," *Proceedings of the IRE*, vol. 45, pp. 1535-1537, 1957.

[4] W. Snodgrass, W. Hafez, N. Harff, and M. Feng, "Pseudomorphic InP/InGaAs heterojunction bipolar transistors (PHBTs) experimentally demonstrating fT = 765 GHz at 25 °C increasing to fT = 845 GHz at -55 °C," in *Proceedings of the 2006 IEEE International Electron Devices Meeting*, 2006, pp. 22.1.1-22.1.4.

[5] T. S. Low, M. W. Dvorak, M. Farhoud, R. E. Yeats, M. Iwamoto, G. K. Essilfie, T. Engel, B. Keppeler, J. S. C. Chang, J. Hadley, G. Patterson, F. Kellert, N. Moll, S. R. Bahl, C. P. Hutchinson, E. Ehlers, M. E. Adamski, M. K. Culver, D. D'Avanzo, and T. Shirley, "GaAsSb DHBT IC technology for RF and microwave instrumentation," in *Proceedings of the IEEE Compound Semiconductor Integrated Circuit Symposium*, 2005, pp. 69-72.

[6] W. Snodgrass, B.-R. Wu, W. Hafez, K. Y. Cheng, and M. Feng, "Performance enhancement of composition-graded-base type-II InP/GaAsSb double-heterojunction bipolar transistors with fT > 500 GHz," *Applied Physics Letters*, vol. 88, p. 222101, 2006.

[7] R. Lövblom, R. Flückiger, Y. Zeng, O. Ostinelli, A. R. Alt, H. Benedickter, and C. R. Bolognesi, "InP/GaAsSb DHBTs with 500-GHz maximum oscillation frequency," *IEEE Electron Device Letters*, vol. 32, pp. 629-631, 2011.

[8] K. Y. (Donald) Cheng, C. C. Liao, H. Xu, K. Y. (Norman) Cheng, and M. Feng, "Hot electron injection effect on the microwave performance of Type-I/II AlInP/GaAsSb/InP double heterojunction bipolar transistors," *Applied Physics Letters*, vol. 98, p. 242103, 2011.

[9] A. W. Hanson, S. A. Stockman, and G. E. Stillman, "Comparison of $In_{0.5}Ga_{0.5}P$/GaAs single-and double-heterojunction bipolar transistors with a carbon-doped base," *IEEE Electron Device Letters*, vol. 14, pp. 25-27, 1993.

[10] B. R. Wu, M. W. Dvorak, P. Colbus, T. S. Low, and D. D'Avanzo, "Performance improvement of composition-graded AlGaAsSb/InP double heterojunction bipolar transistors," in *Proceedings of the International Conference on Indium Phosphide and Related Materials*, 2009, pp. 20-23.

[11] J. G. Ruch, "Electron dynamics in short channel field effect transistors," *IEEE Electron Device Letters*, vol. 19, pp. 652-654, 1972.

[12] T. Ishibashi, "Nonequilibrium electron transport in HBTs," *IEEE Transactions on Electron Devices*, vol. 48, pp. 2595-2604, 2001.

[13] L. H. Camnitz and N. Moll, "An analysis of the cutoff-frequency behavior of microwave heterojunction bipolar transistors," in *Compound Semiconductor Transistors*. Piscataway, NJ: IEEE Press, 1992, pp. 21-45.

[14] J. R. Hayes, A. F. J. Levi, A. C. Gossard, and J. H. English, "Base transport dynamics in a heterojunction bipolar transistor," *Applied Physics Letters*, vol. 49, pp. 1481-1483, 1986.

[15] C. T. Kirk, "A theory of transistor cutoff frequency (fT) falloff at high current densities," *IRE Transactions on Electron Devices*, vol. ED-9, pp. 164-174, 1962.

[16] W. Hafez, J.-W. Lai, and M. Feng, "Vertical scaling of 0.25 μm emitter InP/InGaAs single heterojunction bipolar transistors with f_T of 452 GHz," *IEEE Electron Device Letters*, vol. 24, pp. 438-439, 2003.

[17] W. Hafez, W. Snodgrass, and M. Feng, "12.5 nm base pseudomorphic heterojunction bipolar transistors achieving fT = 710 GHz and fMAX = 340 GHz," *Applied Physics Letters*, vol. 87, p. 252109, 2005.

[18] C. R. Bolognesi, N. Matine, M. W. Dvorak, P. Yeo, X. G. Xu, and S. P. Watkins, "InP/GaAsSb/InP double HBTs: A new alternative for InP-based DHBTs," *IEEE Transactions on Electron Devices*, vol. 48, pp. 2631-2639, 2001.

[19] M. W. Dvorak, C. R. Bolognesi, O. J. Pitts, and S. P. Watkins, "300 GHz InP/GaAsSb/InP double HBTs with high current capability and BV_{CEO} > 6 V," *IEEE Electron Device Letters*, vol. 22, pp. 361-363, 2001.

[20] H. G. Liu, S. P. Watkins, and C. R. Bolognesi, "15-nm base type-II InP/GaAsSb/InP DHBTs with FT=384 GHz and a 6-V BVCEO," *IEEE Transactions on Electron Devices*, vol. 53, pp. 559-561, 2006.

[21] W. Snodgrass, B. R. Wu, K. Y. Cheng and M. Feng, "Type-II GaAsSb/InP DHBT with record f_T = 670 GHz and simultaneous f_T, f_{MAX} > 400 Ghz," in *Proceedings of the 2007 IEEE International Electron Devices Meeting*, 2007, pp. 663-666.

[22] H. G. Liu, O. Ostinelli, Y. Zeng, and C. R. Bolognesi, "600 GHz InP/GaAsSb/InP DHBTs grown by MOCVD with a Ga(As,Sb) graded-base and fTxBVCEO 2.5 THz-V at room temperature," in *Proceedings of the 2007 IEEE International Electron Devices Meeting*, 2007, pp. 667-670.

[23] H. G. Liu, O. Ostinelli, Y. P. Zeng, and C. R. Bolognesi, "High-current-gain InP/GaInP/GaAsSb/InP DHBTs with fT = 436 GHz," *IEEE Electron Device Letters*, vol. 28, pp. 852-855, 2007.

[24] S. W. Cho, J. H. Yun, D. H. Jun, J. I. Song, I. Adesida, N. Pan, and J. H. Jang, "High performance InP/InAlAs/GaAsSb/InP double heterojunction bipolar transistors," *Solid-State Electronics*, vol. 50, pp. 902-907, 2006.

[25] Y. Oda, K. Kurishima, N. Watanabe, M. Uchida, and T. Kobayashi, "High-current-gain InAlP/AlGaAsSb/InP HBTs with a compositionally-graded AlGaAsSb base grown by MOCVD," in *Proceedings of the 2006 International Conference on Indium Phosphide and Related Materials*, 2006, pp. 92-95.

[26] M. Zaknoune, H. Colder, D. A. Yarekha, G. Dambrine, and F. Mollot, "Current gain enhancement in GaAsSb/InP - DHBT type grown by MBE with a graded composition AlInP emitter," in *Proceedings of the 20th International Conference on Indium Phosphide and Related Materials*, 2008, pp. 1-3.

[27] C. G. Van de Walle, "Band lineups and deformation potentials in the model-solid theory," *Physical Review B*, vol. 39, p. 1871, 1989.

[28] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, " Band parameters for III–V compound semiconductors and their alloys," *Journal of Applied Physics*, vol. 89, pp. 5815-5875, 2001.

[29] G. K. Reeves and H. B. Harrison, "Obtaining the specific contact resistance from transmission line model measurements," *IEEE Electron Device Letters*, vol. 3, pp. 111-113, 1982.

[30] W. Liu, *Handbook of III-V Heterojunction Bipolar Transistors*. New York, NY: Wiley-Interscience, 1998.

[31] S. J. Spiegel, D. Ritter, R. A. Hamm, A. Feygenson, and P. R. Smith, "Extraction of the InP/GaInAs heterojunction bipolar transistor small-signal equivalent circuit," *IEEE Transactions on Electron Devices*, vol. 42, pp. 1059-1064, 1995.